

CARTILLA PARA LA ENSEÑANZA DE
LA ESTADÍSTICA EN EL GRADO 11°

HOLMAN HERNANDO DUSSÁN DUSSÁN

Código 2006135058

Universidad Surcolombiana

Facultad de Educación

Licenciatura en Matemáticas

Neiva (Huila)

2011

Índice general

Agradecimientos	3
Introducción	4
1. INICIACIÓN A LA ESTADÍSTICA	6
1.1. Historia de la Estadística	6
1.1.1. Los orígenes de la Estadística en Colombia	8
1.2. ¿Qué es la Estadística? y sus divisiones	10
1.2.1. Estadística descriptiva	12
1.2.2. Estadística inferencial	12
1.3. Aplicaciones de la Estadística	12
1.4. Definición de términos	14
1.4.1. Población	14
1.4.2. Muestra	14
1.4.3. Variables cualitativas	15
1.4.4. Variables cuantitativas	16
2. TABLAS Y GRÁFICAS	17
2.1. Recuento y agrupación de datos	20
2.2. Tablas de frecuencias	21
2.2.1. Tipos de frecuencias	21
2.3. Gráficas	33

2.3.1.	Diagrama de barras	35
2.3.2.	Diagramas circulares	37
2.3.3.	Pictogramas	38
2.3.4.	Cartogramas	40
2.3.5.	Histograma de frecuencias	40
2.3.6.	Polígono de frecuencias	41
2.4.	Aplicaciones en Excel	43
2.4.1.	Construcción de tablas de frecuencias para variable discreta	43
2.4.2.	Construcción de tablas de frecuencias para variable continua (intervalos)	49
2.4.3.	Construcción de gráficos en Excel	61
2.5.	Talleres de aplicaciones	77
3.	MEDIDAS ESTADÍSTICAS	81
3.1.	Medidas de Tendencia central para datos no agrupados	81
3.1.1.	Media aritmética	82
3.1.2.	Mediana	84
3.1.3.	Moda	86
3.2.	Medidas de posición para datos no agrupados	88
3.2.1.	Cuartiles	89
3.2.2.	Deciles	89
3.2.3.	Percentiles	90
3.3.	Medidas de dispersión para datos no agrupados	92
3.3.1.	Rango	93
3.3.2.	Varianza	94
3.3.3.	Desviación estándar	94
3.3.4.	Coefficiente de variación	95
3.4.	Medidas de tendencia central para datos agrupados	100
3.4.1.	Media aritmética	101
3.4.2.	Mediana	103

3.4.3. Moda	104
3.5. Medidas de posición para datos agrupados	104
3.5.1. Cuartil, Decil y Percentil	104
3.6. Medidas de dispersión para datos agrupados	108
3.6.1. Varianza	108
3.6.2. Desviación estándar	109
3.6.3. Coeficiente de variación	109
3.7. Talleres de aplicaciones	115
4. EXPERIMENTOS ALEATORIOS Y PROBABILIDAD	118
4.1. Reseña histórica de la Probabilidad	122
4.1.1. Operaciones entre eventos y probabilidad	126
4.2. Probabilidad condicionada	129
4.3. Técnicas de conteo	131
4.3.1. Principio de multiplicación	131
4.3.2. Permutaciones	133
4.3.3. Combinaciones	137
4.4. Talleres de aplicaciones	140
5. PRUEBAS SABER 11°	143

Agradecimientos

Al lograr esta meta, quiero hacer llegar el más profundo agradecimiento a todas aquellas personas que de una u otra forma han participado en mi preparación profesional brindándome conocimientos, consejos, motivación, apoyo y confianza, por ello, hoy también han triunfado ustedes.

A mis padres

Gracias por darme la vida, por haberme querido tanto, por inculcarme el camino correcto, por haberme hecho una persona de bien, por enseñarme buenos principios. Gracias Dios por haberme dado los mejores padres del mundo, gracias a ellos porque me tomaron de la mano para enseñarme a caminar.

A la profesora Sonia Sierra Cardozo

Por ayudarme, por aportarme sus conocimientos, por corregirme los errores con profesionalismo, por comprenderme y apoyarme, gracias por su valiosa colaboración.

A mis amigos con quienes recorrí esta etapa de mi vida, quienes me enseñaron lo valioso de su amistad, en especial a Paola Andrea Gómez Cedeño.

Agradezco a todo el cuerpo docente del Programa de Licenciatura en Matemáticas porque siempre estuvieron inculcándome conocimientos, ética, profesionalismo y sobre todo por tener la paciencia necesaria en esta carrera.

A todos ustedes cuyos nombres no aparecen en estas líneas, pero que siempre estuvieron y estarán en mi preparación profesional.

Introducción

La estadística es una de las principales ramas de la matemática y se encuentra entre las más estudiadas desde la antigüedad hasta el presente. Además ha jugado un rol fundamental en el desarrollo de otras áreas de la matemática y del conocimiento y hoy en día se utiliza ampliamente en distintas actividades diarias.

El trabajo se divide en cinco capítulos; en el primero, se enmarca la historia, las divisiones y los principales terminos de la estadística. En el segundo capítulo se ofrece un estudio de las tablas de frecuencia y las principales gráficas existentes para representar los resultados de un estudio estadístico, adicionalmente se articulan estos dos temas con la informática, para ser más preciso en el manejo de Microsoft Office Excel 2007.

El tercer capítulo está dedicado a las medidas estadísticas, es decir, las medidas de tendencia central, de posición y de dispersión tanto para datos no agrupados, como para los agrupados en tablas de frecuencias. El cuarto capítulo está dedicado a los experimentos aleatorios y a todo lo referente a la probabilidad matemática.

El último capítulo se encuentra integrado por preguntas tipo ICFES, en las cuales se aplican los temas desarrollados en los anteriores capítulos, buscando una preparación integral para presentar las pruebas SABER 11.

INICIACIÓN A LA ESTADÍSTICA

Es común encontrarse con la Estadística en la cotidianidad y en la mayoría de ocasiones ésta juega un papel fundamental en la interpretación del mundo moderno, aunque a veces no se nota muy claramente su presencia.

Más aún, tal vez ha surgido la inquietud por conocer de donde surge la Estadística como tal, sus orígenes, el por qué se utilizó esta palabra y no otra.

Cabe resaltar en primera instancia que la Estadística no surgió de un momento para otro, sino mediante un proceso largo de desarrollo y evolución; desde hechos de simple recolección de datos hasta la diversa y rigurosa interpretación de estos.

Por eso es conveniente hacer un nutrido viaje que involucre a las principales civilizaciones de la humanidad a través de la HISTORIA DE LA ESTADÍSTICA:

1.1. Historia de la Estadística

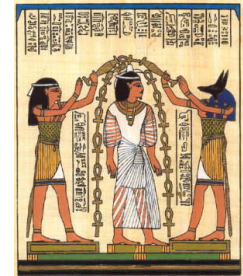
Desde sus orígenes, el ser humano ha buscado la manera de organizarse para tener un mejor control de sus actividades y de los resultados de ellas.

Se sabe que una forma sencilla de estadística se aplicó aproximadamente hace 5000 años, y consistía en contar personas, animales o cosas que se presentaban gráficamente con símbolos en pieles, rocas, palos de madera y paredes de cuevas.



Culturas como la china y griega, entre otras realizaban censos en sus territorios con el fin de reunir datos en torno a la agricultura, productividad, cobranza de impuestos, derechos de votos y servicio militar.

- * Se sabe que los egipcios, analizaban datos de población y riquezas mucho antes de construir sus famosas pirámides. Dichos registros se realizaron con el objetivo de preparar la construcción de las mismas. Llevaban la cuenta de los movimientos poblacionales y continuamente hacían censos por órdenes del faraón. Tal era su dedicación por llevar siempre una relación de todo que hasta tenían a la diosa **Safnkit**, diosa de los libros y las cuentas.



- * Incluso la Biblia da referencias estadísticas, pues es de mencionar, que según el Nuevo Testamento, Jesús nació en Belén porque sus padres debieron acudir allí por un censo ordenado por el emperador romano Augusto. Otros libros bíblicos como el de “**Los Números**” dan referencias de censos en el antiguo pueblo israelí.



- * El Imperio romano fue el primer gobierno que recopiló una gran cantidad de datos sobre la población, superficie y renta de todos los territorios bajo su control. Cada cinco años realizaban un censo “Census” de la población y sus funcionarios públicos tenían la obligación de anotar nacimientos, defunciones y matrimonios, sin olvidar los recuentos periódicos del ganado y de las riquezas contenidas en las tierras conquistadas.
- * La recopilación sistemática de datos económicos y de población se inició en Venecia

y Florencia (ciudades italianas), durante el Renacimiento. La idea de recopilación y obtención de datos de interés estatal se extendió desde Italia a otros países de Europa occidental. De hecho, durante la primera mitad del siglo XVI, era habitual que los gobiernos europeos obligaran a las parroquias a que registraran los nacimientos, matrimonios y defunciones. Debido a las muy escasas condiciones de salud pública, las estadísticas referidas a estos hechos tenían un especial interés.

✱ En 1662 el comerciante inglés John Graunt publicó un libro titulado “**Observaciones naturales y políticas hechas a partir de los datos de mortalidad**”.

Los trabajos de Graunt, sobre tablas de mortalidad inspiraron el aporte hecho por Edmund Halley en 1693. Halley, descubridor del cometa que lleva su nombre, utilizó las tablas de mortalidad para calcular la probabilidad que una persona de cualquier edad tenía de sobrepasar otra edad distinta. Halley ejerció gran influencia para convencer a las aseguradoras del momento de que los seguros tenían que depender de las edades de los asegurados.

La recopilación de datos se incrementó de manera continua durante todo el siglo XVII y hasta bien entrado el XVIII. Por ejemplo, la ciudad de París empezó a registrar cifras de mortalidad en 1667; y en 1730 registrar las edades de muerte, era una práctica común en toda Europa.

1.1.1. Los orígenes de la Estadística en Colombia

Los historiadores colombianos Enrique Otero D’Costa y Gerardo Arrubla afirman que “de los tiempos anteriores a la conquista no se conocen datos por los cuales se pueda afirmar que los primitivos habitantes de Colombia hubiesen desarrollado actividad alguna estadística”.

Con referencia a los chibchas, a los que denominan “los más civilizados”, dicen que: “apenas si podían contar hasta la docena y muy difícilmente hasta el centenar”.

No obstante, se habla de que tenían una economía de trueque y usaban como signo de cambio una moneda de oro, llamada “cuca”.



Pero lo que denota más la presencia de una actividad colindante con la estadística, ya no de carácter indirecto sino manifiesto, es el establecimiento de una contabilidad incluso con anotación de ventas a crédito e intereses. Ello se explica porque su intercambio comercial se había extendido hasta las tribus ribereñas del alto, medio y bajo Magdalena, con establecimiento de mercados chibchas en Sorocotá, Neiva y Santa Marta.

En la colonia lo que pudiera llamarse estadística se concreta a los relevamientos para reparto de las encomiendas, número de esclavos, reclutamientos, impuestos y contribuciones. De 1659 en adelante existen documentos que revelan la existencia de estadísticas sobre actividad civil, militar, eclesiástica, administrativa y de gobierno en general. Se registraban así mismo las operaciones de navegación marítima, fluvial y terrestre, detallada y ordenadamente. Las fuentes estaban constituidas por la nutrida correspondencia de las autoridades, generalmente militares, dirigida al Virrey.

A partir del recuento de población de 1770, el período intercensal con el siguiente es de 8 años; de este con el que se levanta inmediatamente después, de solo 4 años; viene luego un interregno de 21 años y finalmente torna a reducirse el plazo intercensal, esta vez a 7 años, con el censo de 1810, ya en plena efervescencia revolucionaria.

En 1783, se inicia la Expedición Botánica, impulsada con el propósito y ante las urgentes necesidades de la Corona, en ruina económica, de descorrer el velo de las riquezas del trópico, para su beneficio, ya que la Nueva Granada solo había proporcionado oro y plata a la metrópoli. La Expedición es también motivo de nuevo impulso estadístico, el que se patentiza en trabajos tales como la “Relación Territorial de la Provincia de Pamplona”, de Joaquín Camacho; la “Razón de Nativos, Casados y Muertos en Popayán (1800-1804)”, y el “Estado del Hospital Real de Popayán”, ambos de Francisco Mosquera la “Estadística de Santafé”,

de Nicolás Francisco Omaña; la “Estadística de Nacidos y Muertos en Puente Real (1800-1808)”, de José María Durán, y otros estudios estadísticos sobre meteorología, cultivos, etc.

La Ley de 1825 (marzo 11) dispuso el levantamiento del censo de población. Un mes después, el 25 de abril, se establece por decreto la Oficina Central de Estadística, encargada de llevar a la práctica la operación censal. Esta entidad quedó bajo la dependencia de la Secretaría del Despacho del Interior de la Gran Colombia (Nueva Granada, Quito y Venezuela), formada por la Ley de 1819 (diciembre 17).

En octubre de 1951, mediante Decreto 2240, se crea la Dirección Nacional de Estadística, dependencia directa de la Presidencia de la República.

Dos años después, bajo el gobierno del general Gustavo Rojas Pinilla, y mediante Decreto 2666, nace el Departamento Administrativo Nacional de Estadística DANE, cuyo objetivo era garantizar la producción, disponibilidad y calidad de la información estadística estratégica, y dirigir, planear, ejecutar, coordinar, regular y evaluar la producción y difusión de información oficial básica.



En la actualidad es la encargada de la elaboración de los censos a nivel nacional, así como arrojar los datos estadísticos en torno a crecimiento económico, demográfico y de proyección nacional.

1.2. ¿Qué es la Estadística? y sus divisiones

La Estadística es mucho más que sólo números apilados y gráficas bonitas. Godofredo Achenwall, estadístico alemán, profesor de la Universidad de Gotinga, acuñó en 1760 la palabra ESTADÍSTICA, derivada del vocablo italiano statista (estadista), a su vez proveniente del latín status (estado o situación).

En su origen, por tanto, la Estadística estuvo asociada a los estados, para ser utilizados

por el gobierno y cuerpos administrativos (a menudo centralizados). La recolección de datos acerca de estados y localidades continúa ampliamente a través de los servicios de estadística nacionales e internacionales. En particular, los censos suministran información regular acerca de la población.

Comúnmente es considerada como una colección de hechos numéricos expresados en términos de una relación sumisa, y que han sido recopilados a partir de datos numéricos.

Definen la estadística como un valor resumido, calculado, como base en una muestra de observaciones que generalmente, aunque no por necesidad, se considera como una estimación de parámetro de determinada población; es decir, una función de valores de muestra.



Formalmente, la estadística es una ciencia que tiene como finalidad facilitar la solución de problemas en los cuales se necesita conocer algunas características sobre el comportamiento de algún suceso o evento. Características que permiten conocer o mejorar el conocimiento de ese suceso. Además permiten inferir el comportamiento de sucesos iguales o similares sin que estos ocurran.

Como ciencia se encarga de recopilar, e interpretar datos que en el futuro servirán para proyectar posibles problemáticas futuras, consiguiendo según estos datos, la solución más viable y rápida.

El objetivo básico de la estadística es hacer inferencia acerca de una población con base en la información obtenida a partir de una muestra, ¿qué significa esto? inferir significa inducir una cosa de otra, llevar consigo, conducir a un resultado.

Es decir, se pretende establecer inferencia acerca de una población, entendiendo a la población como un conjunto de individuos, organismos o entes inanimados de los cuales se quiere conocer alguna o algunas características para que ayuden a tomar una decisión u obtener alguna conclusión de mucha importancia.

La Estadística se divide en dos grandes ramas:

1.2.1. Estadística descriptiva

Se dedica a los métodos de recolección, descripción, visualización y resumen de los datos organizados a partir del estudio de un fenómeno. Los datos se pueden recopilar, estructurar y analizar, numérica y gráficamente.

Aunque hay tendencia a generalizar a toda la población, las primeras conclusiones obtenidas tras un análisis descriptivo, es un estudio calculando una serie de medidas de tendencia central, para ver en qué forma los datos se agrupan o dispersan en torno a un valor central.

1.2.2. Estadística inferencial

Se dedica a describir los datos y además, a deducir conclusiones que permitan realizar el estudio de un conjunto más amplio de datos.

La estadística inferencial, es el proceso por el cual se deducen (infieren) propiedades o características de una población a partir de una muestra significativa. Uno de los aspectos principales de la inferencia es la estimación de parámetros estadísticos.

1.3. Aplicaciones de la Estadística

Dos de las áreas más importantes de aplicación de la estadística a principios del siglo XX son la biología y la agricultura; todo ello gracias a los logros del científico inglés Ronald A. Fisher. La Teoría de la Inferencia desarrollada por estos investigadores pioneros y otros entre los que se encuentran Egon Pearson, hijo de Karl Pearson y el estadístico matemático polaco Jerzy Neyman fue lo suficientemente general como para tratar una amplia gama de problemas cuantitativos prácticos. Como resultado, tras los primeros años del siglo XX, aumentó rápidamente el número de personas, procedentes de la ciencia, los negocios y la

administración que empezaron a considerar la estadística como una herramienta capaz de suministrar soluciones cuantitativas a una gran variedad de problemas científicos y prácticos.

Aunque comúnmente se asocie a estudios demográficos, económicos y sociológicos, gran parte de los logros de la estadística se derivan del interés de los científicos por desarrollar modelos que expliquen el comportamiento de las propiedades de la materia y de los caracteres biológicos. La Medicina, la Biología, la Física y en definitiva, casi en todos los campos de las ciencias emplean instrumentos estadísticos de importancia fundamental para el desarrollo de sus modelos de trabajo.

En las ciencias naturales, se emplea en la descripción de modelos termodinámicos complejos (mecánica estadística), en física cuántica, en mecánica de fluidos o en la teoría cinética de los gases, entre muchos otros campos.

En las ciencias sociales y económicas, es un pilar básico del desarrollo de la demografía y la sociología aplicada. Cabe aclarar que la demografía es el estudio estadístico de una colectividad humana, referido a un determinado momento o a su evolución.

En economía, suministra los valores que ayudan a descubrir interrelaciones entre múltiples parámetros macro y microeconómicos.

En las ciencias médicas, permite establecer pautas sobre la evolución de las enfermedades y los enfermos, los índices de mortalidad, natalidad, el grado de eficacia de un medicamento, etc.



1.4. Definición de términos

1.4.1. Población

La **población**, es el conjunto de todos los individuos (personas, objetos, animales, etc) de los cuales se obtiene información sobre el fenómeno que se estudia.

Por ejemplo, si se estudia el precio de los servicios públicos en una ciudad, la población será el total de las viviendas de dicha ciudad. Evidentemente, es bastante dispendioso recopilar información de todas las viviendas de la ciudad, en cambio, se selecciona un subgrupo (muestra) que sea lo suficientemente representativo.

1.4.2. Muestra

La **muestra** es un subconjunto de una población que usualmente es seleccionada de tal modo que ponga de manifiesto las propiedades de la población y sea representativa

Ejemplo:

Identificar en las siguientes situaciones la población y la muestra:

- a. Una emisora de radio quiere saber cuál es el programa con mayor audiencia en una ciudad de Colombia. Para ello contrata una empresa que lleva a cabo una encuesta telefónica para recoger la información. La empresa, divide a sus operadores por zonas. Los operadores logran encuestar el 70 % de la base de datos que tiene la empresa.
 - La población la conforman todos los habitantes de la ciudad, y la muestra es el 70 % de los habitantes de la ciudad que lograron encuestar.

- b. El Ministerio de Protección Social desea determinar los índices de mortalidad infantil en Medellín. Para ello va a realizar un estudio del número de casos registrados en algunos hospitales.
- En este caso, la población está representada por los niños y niñas de la ciudad de Medellín y la muestra la constituyen los casos registrados en algunos hospitales de la ciudad.
- c. Una agencia de viajes decide realizar una encuesta a sus clientes del Huila a través del correo electrónico donde se les pregunta el destino preferido para pasar sus vacaciones en el departamento del Huila.
- En este caso, la población y la muestra son exactamente iguales y son los clientes de la agencia de viajes.

1.4.3. Variables cualitativas

Las variables **cualitativas**, denominadas también atributos, son todas aquellas que se puedan describir mediante palabras.

Ejemplos:

- ✗ Los empleados de una empresa clasificados por cargos.
 - ✗ Los nombres de las sucursales de un almacén.
 - ✗ Las marcas de aceite de uso doméstico.
 - ✗ Los días de la semana.
 - ✗ Los departamentos de Colombia.
 - ✗ El color de los ojos.
-

1.4.4. Variables cuantitativas

Las variables **cuantitativas**, son aquellas susceptibles de ser expresadas numéricamente.

Ejemplos:

- ❖ El peso de una persona.
- ❖ La cantidad de elementos comprados en una tienda.
- ❖ El sueldo mensual de un empleado.
- ❖ La edad de una persona.
- ❖ El número de hijos en una familia.

Las variables cuantitativas se clasifican en discretas y continuas.

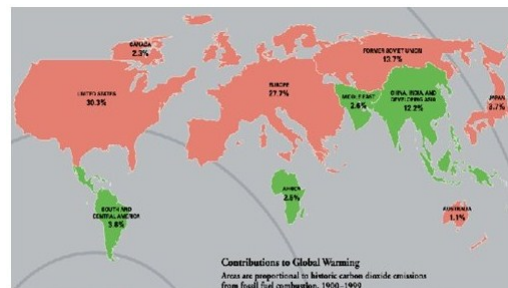
La variable cuantitativa discreta toma valores específicos, por lo general son números enteros. Por ejemplo, cantidad de hermanos, cantidad de colegios.

La variable cuantitativa continua toma valores que pertenecen a un intervalo de números reales. Por ejemplo, estatura, velocidad, distancia, peso, tiempo. Además, las expresadas en medidas de relación, tales como porcentajes, tasas.

TABLAS Y GRÁFICAS

EL CALENTAMIENTO GLOBAL

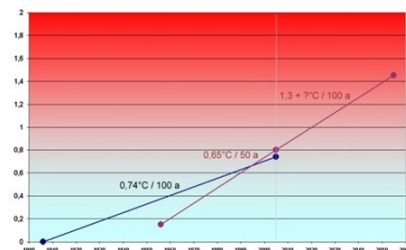
El calentamiento global es un término utilizado para referirse al cambio climático; fenómeno debido al aumento de la temperatura media global de la atmósfera terrestre y de los océanos, este aumento en la temperatura se observa en registros que se vienen haciendo desde el año 1850. Este incremento se habría acentuado en las últimas décadas del siglo XX y en la primera del XXI.



Mapa del World Resources Institute muestra el tamaño de los países y continentes proporcionales a su contribución al calentamiento global en 2002.

La principal consecuencia debido al calentamiento global es el efecto invernadero.

El aumento de la temperatura no es constante, se está acelerando. Si se toman los valores medidos de los últimos 50 años como una base para el aumento de 1956 a 2055 terminaríamos en un aumento medio de 1.3°C por 100 años.



Se señala que si se derritiera únicamente la capa de hielo de Groenlandia, el nivel del mar podría subir siete metros, dejando bajo el agua a muchas ciudades.

En un estudio los científicos de la Universidad de Bristol (Reino Unido), han calculado que si el incremento sobrepasa los dos grados, ya se notarán las consecuencias en el medio ambiente, entre las cuales se presentarían más incendios forestales, más deforestación y más inundaciones. Pero si la temperatura supera los tres grados, se esperarán catástrofes incontrollables.

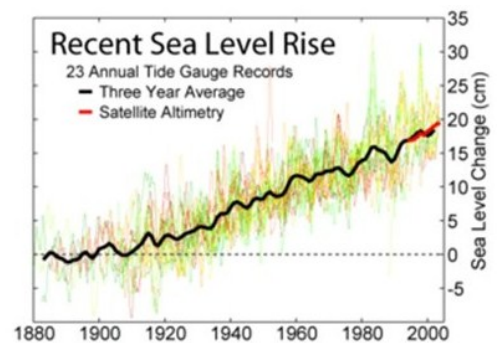
Las zonas más afectadas ante un aumento de la temperatura del planeta serían la Amazonia y las regiones árticas. La Amazonia sufriría de sequía y en cuanto a las regiones árticas, los bosques boreales y la tundra del Ártico desaparecerían.

Si la temperatura aumenta 2 o 3 grados, el 30 % de los bosques podrán estar en peligro, y el 60 % si supera los 3 grados.

Los periodos de sequía serán cada vez más intensos en la zona oeste del continente africano, al este de EEUU, América central y países del sur de Europa.

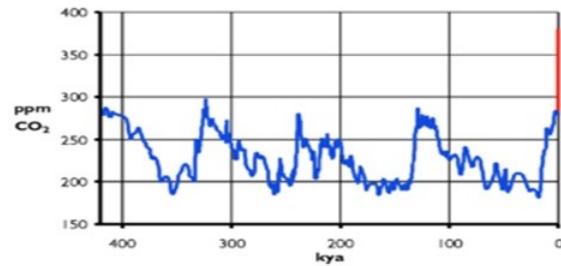
Un informe del Centro de Seguridad Nacional de USA advierte que: “en los próximos 30 o 40 años uno de los motivos de las guerras será el agua y por ende una creciente inestabilidad causada por hambruna, por enfermedades y olas de refugiados y desplazados”.

Otro motivo de gran preocupación es la elevación del nivel de los mares. Los niveles de los mares se están elevando entre 1 y 2 centímetros por decenio, a la vez que se agudizan los fenómenos climáticos extremos, y algunas naciones isleñas del Océano Pacífico, como Tuvalu, ya están trabajando en los detalles de una eventual evacuación.



El calentamiento global modificaría la distribución de la fauna y flora del planeta. Ello supondría la expansión de enfermedades de las que algunos de esos animales son portadores. Tal es el caso de la malaria, el dengue o la fiebre amarilla, cuyos vectores son ciertas especies de mosquitos que habitan principalmente en zonas tropicales.

Las industrias que utilizan el carbón como fuente de energía, los tubos de escape de los automóviles, las chimeneas de las fábricas y otros subproductos gaseosos procedentes de la actividad humana, contribuyen con cerca de 22.000 millones de toneladas de dióxido de carbono y otros gases de efecto invernadero a la atmósfera terrestre cada año.



Concentración de dióxido de carbono en los últimos 417.000 años. La línea roja indica la variación a partir de 1800.

Soluciones domésticas para reducir la emisión de CO_2

Algunas de las soluciones que cada uno de nosotros podemos aplicar para controlar la producción de CO_2 :

- * Cambiar las bombillas tradicionales por otras de bajo consumo.
- * Ajustando el aire acondicionado se podrían ahorrar unos 900 kilos de dióxido de carbono al año.
- * Comprar productos de papel reciclado.
- * Comprar alimentos frescos. Producir comida congelada consume 10 veces más energía.
- * Evitar comprar productos envasados, de esta manera se reduce en un 10% la basura personal.
- * Utilizar menos los aparatos eléctricos. Desconectar a los que no se esté prestando atención en ese momento.
- * Elegir un vehículo de menor consumo. Un vehículo nuevo puede ahorrar 1360 kilos de dióxido de carbono al año.
- * Usar menos el automóvil, caminar o ir en bicicleta.

- * Sembrar árboles. Un solo árbol elimina una tonelada de dióxido de carbono a lo largo de su vida.



¿Qué está haciendo usted para revertir este proceso?

2.1. Recuento y agrupación de datos

La estadística es un conjunto de técnicas o métodos que permiten la observación, recopilación, ordenación, descripción, y análisis de un fenómeno. La elaboración de tablas o cuadros, facilita el análisis y la presentación de la información.

Para elaborar tablas de frecuencias, se debe, en primera instancia, identificar las características que se investigaron, lo cual permite una mejor clasificación de lo observado.

Cuando se analiza o se estudia una sola característica de los elementos seleccionados en la investigación, se dice que la variable corresponde a una distribución unidimensional.

Cuando se trabajan simultáneamente con dos variables, se trata de distribuciones bidimensionales.

En el caso de estudiar tres o más características simultáneamente, se habla de una distribución pluridimensional o multidimensional; tal sería el caso de clasificar las ventas de una empresa por sucursales, valor y semestres.

2.2. Tablas de frecuencias

Tablas de frecuencias: Una tabla de frecuencia está conformada por los valores que toma la variable y sus correspondientes frecuencias en donde la frecuencia representa el número de veces que un valor determinado se repite. Estas tablas también reciben el nombre distribuciones de frecuencias.

A partir de este momento el estudio se concentra en una sola variable, “estadísticas unidimensionales”.

Las tablas de frecuencias se clasifican según el tipo de variable:

- Tablas de frecuencias para variable cualitativa.
- Tablas de frecuencias para variable cuantitativa.

Una tabla de frecuencias tiene como finalidad presentar en forma ordenada los valores que toman las diferentes variables, de tal forma que permitan al lector tener una visión de un conjunto de datos, aclarando el texto del informe o complementándolo. Bajo este principio los datos se clasifican y ordenan de acuerdo a ciertas características cualitativas y/o cuantitativas, indicándose el número de veces que se repite.

Una **distribución de frecuencias** es el registro de todos los valores de la variable, con sus respectivas frecuencias.

2.2.1. Tipos de frecuencias

A continuación, se describe cada uno de los diferentes tipos de frecuencias asociadas a una distribución.

Frecuencia absoluta (f): es el número de veces que se repite una de las categorías o los valores que toma la variable.

Frecuencia relativa (f_r): se define como la frecuencia absoluta (f) dividida entre el total de observaciones (n). Es decir,

$$f_r = \frac{f}{n}$$

Frecuencia porcentual (%) se obtiene al multiplicar la frecuencia relativa (f_r) por 100.

Las frecuencias acumuladas permiten conocer rápidamente el número de observaciones que están por debajo de una categoría o valor.

Frecuencia absoluta acumulada (F) para calcularla se hace por medio de sumas sucesivas, las que se simbolizan por F . La columna F_1 no se suma, y la última frecuencia absoluta acumulada será igual a n .

Luego:

$$F_1 = f_1$$

$$F_2 = f_1 + f_2 = F_1 + f_2$$

$$F_3 = f_1 + f_2 + f_3 = F_2 + f_3; \text{ y así sucesivamente.}$$

⋮

$$F_i = f_1 + f_2 + f_3 + f_4 + \dots + f_i = \sum_{j=1}^i f_j$$

Frecuencias relativas acumuladas (F_r), para calcularla se procede al igual que en las frecuencias absolutas acumuladas.

Luego:

$$F_{r1} = f_{r1}$$

$$F_{r2} = f_{r1} + f_{r2} = F_{r1} + f_{r2}$$

$F_{r3} = f_{r1} + f_{r2} + f_{r3} = F_{r2} + f_{r3}$; y así sucesivamente.

⋮

$$F_{ri} = f_{r1} + f_{r2} + f_{r3} + f_{r4} + \dots + f_{ri} = \sum_{j=1}^i f_{rj}$$

Tablas de frecuencias para variable cualitativa:

Ejemplo:

A un grupo de estudiantes se les pregunto por su color preferido. Los resultados fueron los siguientes: rojo, azul, verde, rojo, rojo, azul, verde, verde, verde, azul, rojo, verde, azul, azul, verde, rojo, rojo y rojo.

Luego, al organizar los datos en la tabla de frecuencias se tiene:

Variable frecuencia

Color preferido	f
Rojo	7
Azul	5
Verde	6
Total	18

Ejemplo:

El rector del colegio “Huila, Tierra de Promisión”, le pidió a los estudiantes del grado 11°, que realizaran un estudio para saber la ubicación por carreras profesionales de los egresados de la última promoción del colegio.

Los resultados que obtuvieron fueron los siguientes:

Medicina	Enfermería	Psicología	Medicina
Derecho	Ingeniería civil	Psicología	Ingeniería de Sistemas
Ingeniería de Sistemas	Medicina	Enfermería	Psicología
Derecho	Derecho	Comunicación Social	Ingeniería de Sistemas
Comunicación Social	Medicina	Enfermería	Ingeniería Civil

La forma más práctica, fácil y entendible que tiene los estudiantes de grado 11° para suministrarle el resultado del estudio investigativo al rector del colegio es elaborando una distribución de frecuencias, como la siguiente:

Carrera	f	f_r	F	F_r	%
Medicina	4	$\frac{4}{20} = 0,20$	4	0,20	$0,20 \times 100 = 20\%$
Enfermería	3	$\frac{3}{20} = 0,15$	$4 + 3 = 7$	$0,20 + 0,15 = 0,35$	$0,20 \times 100 = 20\%$
Psicología	3	$\frac{3}{20} = 0,15$	$7 + 3 = 10$	$0,35 + 0,15 = 0,50$	$0,15 \times 100 = 15\%$
Derecho	3	$\frac{3}{20} = 0,15$	$10 + 3 = 13$	$0,50 + 0,15 = 0,65$	$0,15 \times 100 = 15\%$
C. Social	2	$\frac{2}{20} = 0,10$	$13 + 2 = 15$	$0,65 + 0,10 = 0,75$	$0,10 \times 100 = 10\%$
Ing. civil	2	$\frac{2}{20} = 0,10$	$15 + 2 = 17$	$0,75 + 0,10 = 0,85$	$0,10 \times 100 = 10\%$
Ing. de Sistemas	3	$\frac{3}{20} = 0,15$	$17 + 3 = 20$	$0,85 + 0,15 = 1,00$	$0,15 \times 100 = 15\%$
Total	20	1			

Tablas de frecuencias para variable cuantitativa discreta:

Ejemplo :

A los estudiantes del grado 11° de la jornada de la mañana del colegio “Huila, Tierra de Promisión”, se les preguntó por el número de veces que ha repetido un año y los resultados obtenidos fueron los siguientes:

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

La distribución de frecuencias es la siguiente:

N° de veces que ha repetido un año	f	F	f_r	F_r	%
0	3	3	0,107	0,107	11 %
1	9	12	0,321	0,428	32 %
2	13	25	0,464	0,892	46 %
3	3	28	0,107	0,999 \approx 1	11 %
Total	28				100 %

Ejemplo:

Al preguntar por el número de niños que tiene cada una de las 50 familias del barrio “La Gaitana”, se obtuvieron los siguientes resultados.

2, 3, 2, 1, 1, 1, 3, 4, 2, 2, 2, 2, 1, 2, 1, 1, 1, 3, 2, 2, 2, 4
 2, 1, 1, 3, 2, 3, 1, 2, 4, 2, 1, 4, 1, 1, 3, 4, 3, 2, 2, 2, 1, 3
 3, 2, 1, 2, 2, 1

Se puede observar que la variable toma valores enteros comprendidos entre 1 y 4. Luego la distribución de frecuencias quedará de la siguiente manera.

Niños por familia	Número de familias
1	16
2	20
3	9
4	5
Total	50

Distribución de frecuencias con intervalos para variable continua

Ejemplo:

En el pasado mes de junio, una empresa de productos de belleza realizó un estudio para saber cuánto estaría dispuesto a pagar el público por un nuevo producto para el cuidado del cabello; por tal motivo le encargo esta tarea a su personal.



Para realizar dicho estudio, el personal opto por preguntarle a las personas que salen de un salón de belleza, el dinero que estaría dispuesta a pagar por el nuevo producto. Los resultados que se obtuvieron fueron los siguientes:

$$\begin{array}{lllll}
 X_1 = 2000 & X_7 = 2250 & X_{13} = 2350 & X_{19} = 2200 & X_{25} = 2300 \\
 X_2 = 2150 & X_8 = 2300 & X_{14} = 2300 & X_{20} = 2100 & X_{26} = 2150 \\
 X_3 = 2200 & X_9 = 2150 & X_{15} = 2150 & X_{21} = 2250 & X_{27} = 2350 \\
 X_4 = 2100 & X_{10} = 2300 & X_{16} = 2300 & X_{22} = 2000 & X_{28} = 2150 \\
 X_5 = 2400 & X_{11} = 2000 & X_{17} = 2350 & X_{23} = 2400 & X_{29} = 2100 \\
 X_6 = 2100 & X_{12} = 2050 & X_{18} = 2250 & X_{24} = 2150 & X_{30} = 2000
 \end{array}$$

¿Cómo pueden los empleados dar a conocer los resultados del estudio, al gerente de la empresa para que decida el precio de venta al público del nuevo producto?.

En algunas situaciones (como en la anterior) los datos recogidos presentan frecuencias muy pequeñas debido a la gran variedad de los valores. Entonces es útil construir una distribución de frecuencias que permita agrupar los datos por intervalos que tengan la misma longitud. Teniendo en cuenta los datos anteriores se explicará el procedimiento a seguir para la elaboración de dicha tabla.

En la elaboración de la tabla de frecuencias, se deben seguir los siguientes pasos:

- a. Se determinan los valores máximos y mínimos que toma la variable X ;

$$X_{\min} = 2000$$

$$X_{\max} = 2400$$

- b. Se halla el Rango o Recorrido que tiene la variable, el cual consiste en la diferencia entre el valor máximo y el valor mínimo.

$$X_{\max} - X_{\min} = \text{Rango o recorrido}$$

$$\text{Rango} = 2400 - 2000 = 400$$

- c. Se determina el número de intervalos que se utilizarán para agrupar los datos:

Una de las formas de obtener el número de intervalos (m) es aplicando la raíz cuadrada del total de datos n , es decir, \sqrt{n} y aproximar su resultado al entero más cercano.

$$m = \sqrt{n}$$

Aplicando la fórmula a nuestro ejercicio, se obtendrá:

$$m = \sqrt{30}$$

$$m = 5,4772$$

$$m \cong 5$$

Luego en el ejercicio se utilizaran cinco intervalos (5).

En la práctica el número de intervalos se determina atendiendo varios factores. En todo caso, se recomienda que el valor hasta donde sea posible, no sea menor de 5, ni mayor de 16.

- d. Ahora después de determinar el número de intervalos, se debe decidir por el valor de la amplitud (C) para cada intervalo.

Se considera el valor de C constante para todos los intervalos. Dicho valor constante se obtiene aplicando la siguiente fórmula:

$$C = \frac{X_{\text{máx}} - X_{\text{mín}}}{m} = \frac{\text{Rango}}{m}$$

Para el ejercicio se tendrá:

$$\begin{aligned} C &= \frac{X_{\text{máx}} - X_{\text{mín}}}{m} \\ &= \frac{2400 - 2000}{5} \\ &= \frac{400}{5} \\ &= 80 \end{aligned}$$

Para facilitar los cálculos, el valor de C , siempre debe ser un número entero, por tal motivo se debe aproximar por exceso al siguiente número entero.

- e. Ahora ya se puede elaborar la tabla de frecuencias. Primero se inicia con la construcción de los intervalos, se inicia con la determinación del valor de $X_{\text{mín}}$ en el nuevo rango, que para este caso es 2000; el cual se toma como límite inferior del primer intervalo, luego se procede a agregarle el valor de la amplitud para así obtener el límite superior, que será a su vez el límite inferior del segundo intervalo, al que luego se le agrega nuevamente el valor de C para obtener el límite superior del segundo intervalo, y así sucesivamente hasta completar el número de intervalos (m). Los intervalos tienen el límite inferior
-

cerrado ([) y el límite superior abierto ()) excepto en el último intervalo donde ambos límites son cerrados. Entonces la columna que se obtiene con los intervalos es la siguiente:

Precio (s)
[2000 – 2080)
[2080 – 2160)
[2160 – 2240)
[2240 – 2320)
[2320 – 2400]

En cada uno de los anteriores intervalos se tiene un extremo cerrado y un extremo abierto con el fin de que en el intervalo donde se encuentre abierto no se tome el dato pero si se debe tomar en el intervalo cerrado, es decir, en el intervalo [2000 - 2080), el dato 2080 no pertenece a este intervalo pero si al siguiente intervalo [2080 - 2160).

- f. Ahora se elabora la siguiente columna de la tabla, la cual consiste en hallar la marca de clase (Y_i), que es un punto representativo del intervalo. La marca de clase no es más que una forma abreviada de representar un intervalo mediante uno de sus puntos. Para ello se toma como representante, el punto medio del mismo. Luego:

$$y = \frac{\text{Límite inferior} + \text{Límite superior}}{2}$$

Para el ejercicio, veamos como se calculan:

$$y_1 = \frac{2000 + 2080}{2} = 2040$$

$$y_2 = \frac{2080 + 2160}{2} = 2120$$

$$y_3 = \frac{2160 + 2240}{2} = 2200$$

$$y_4 = \frac{2240 + 2320}{2} = 2280$$

$$y_5 = \frac{2320 + 2400}{2} = 2360$$

Precio (s)	y
[2000 – 2080)	2040
[2080 – 2160)	2120
[2160 – 2240)	2200
[2240 – 2320)	2280
[2320 – 2400]	2320

- g. Ahora se procede a realizar el conteo, el cual consiste en ubicar cada dato u observación en el intervalo correspondiente, que es la misma frecuencia absoluta (f), así de este modo se obtiene la siguiente columna de la tabla de frecuencias:

Precio (s)	y	f
[2000 – 2080)	2040	5
[2080 – 2160)	2120	10
[2160 – 2240)	2200	2
[2240 – 2320)	2280	8
[2320 – 2400]	2320	5

Nota: No olvidar que la suma de las frecuencias absolutas es igual al tamaño (n) de la muestra o de la población.

- h. Se halla la frecuencia relativa (f_r), frecuencia relativa acumulada (F_r), frecuencia absoluta acumulada (F), el porcentaje (%) siguiendo los procedimientos para cada caso explicados anteriormente, de este modo se completa la tabla de frecuencias, obteniendo el siguiente resultado:

Precio (s)	y	f	f_r	F	F_r	%
[2000 – 2080)	2040	5	0,166	5	0,166	17%
[2080 – 2160)	2120	10	0,333	15	0,499	33%
[2160 – 2240)	2200	2	0,066	17	0,565	7%
[2240 – 2320)	2280	8	0,266	25	0,831	27%
[2320 – 2400]	2320	5	0,166	30	1,000	16%
Total		30				100%

Ejemplo:

A las primeras 16 personas que salen de un supermercado se les pregunta el número de artículos que compraron y los resultados obtenidos fueron los siguientes:



2, 4, 4, 3, 8, 10, 9, 11, 11, 11, 13, 13, 2, 3, 5, 8, 10, 2, 10, 13, 8, 9, 3, 4, 2

Se pide elaborar una tabla de frecuencias con intervalos y una tabla de frecuencias sin intervalos.

Solución

La tabla de frecuencias con intervalos es la siguiente:

Número de artículos	y	f	f_r	F	F_r	%
$[2 - 4, 2)$	3, 1	10	0, 4	10	0, 4	40 %
$[4, 2 - 6, 4)$	5, 3	1	0, 04	11	0, 44	4 %
$[6, 4 - 8, 6)$	7, 5	3	0, 12	14	0, 56	12 %
$[8, 6 - 10, 8)$	9, 7	5	0, 2	19	0, 76	20 %
$[10, 8 - 13]$	11, 9	6	0, 24	25	1, 00	24 %
Total		25				100 %

La tabla de valores sin intervalos para el número de artículos que compran las personas en un supermercado es la siguiente:

Número de artículos	f	F	f_r	F_r	%
2	4	4	0, 16	0, 16	16 %
3	3	7	0, 12	0, 28	12 %
4	3	10	0, 12	0, 40	12 %
5	1	11	0, 04	0, 44	4 %
8	3	14	0, 12	0, 56	12 %
9	2	16	0, 08	0, 64	8 %
10	3	19	0, 12	0, 76	12 %
11	3	22	0, 12	0, 88	12 %
13	3	25	0, 12	1, 00	12 %
Total	25				100 %

2.3. Gráficas

Una vez ordenados los datos con sus respectivas frecuencias, se tiene una primera información del fenómeno estadístico estudiado, sin embargo, la presentación gráfica de estos datos permite expresar la misma información de una manera más intuitiva y de más fácil comprensión.

La presentación de la información obtenida en una encuesta, se puede realizar de tres formas:

- Textual.
- Tablas de frecuencias.
- Gráfica.

Cuando el informe se hace de forma textual, se hace una presentación simple y limitada, tal como sucede en los artículos de los periódicos, donde los datos aparecen intercalados con los comentarios y con las conclusiones derivadas de las interpretaciones de los mismos periodistas.

Si el informe textual va acompañado de cuadros, estos sirven de ayuda al lector para entender o ampliar el texto. Sin embargo, no es suficiente para la presentación de un buen informe, ya que la lectura del texto implicará un tiempo precioso para entender todo el contenido; ahora, si el lector se remite a las tablas de frecuencias, encontrará unas cifras, que en algunos casos no le aclaran lo suficiente y en otros, no le dicen mucho sobre el problema estudiado, de ahí que sea necesario que el informe vaya acompañado de gráficas, no como un sustituto de las formas anteriores de presentación, sino como la mejor manera de visualizar la información.

Reglas para la elaboración de gráficas

La forma de la gráfica depende en gran parte del gusto de la persona que la elabora; sin embargo, hay ciertos principios generales que se deben tener en cuenta en el logro de una buena gráfica. A continuación se indican algunos de los principios más aceptados:

-
- ☛ La mejor gráfica es la más simple. Las líneas y símbolos usados, deben ser los estrictamente indispensables para una mejor visualización de la información.
 - ☛ Si hay más de dos gráficas (figuras), deben ser numeradas, indicándose con su respectivo nombre.
 - ☛ Toda gráfica debe llevar un título que aclare su contenido. Se dice que debe responder a los siguientes interrogantes: qué, cómo, cuándo y dónde.
 - ☛ Las líneas que llevan escalas, deben dibujarse más gruesas que las demás coordenadas; a su vez, las líneas que representan los datos deben ser más gruesas que las de las escalas.
 - ☛ La línea vertical, denominada ordenada, se utiliza para representar las frecuencias, las cuales deben iniciar en cero.
 - ☛ Las características cualitativas y cuantitativas, por lo general, van en línea horizontal.
 - ☛ La lectura de la escala del eje horizontal se hace de izquierda a derecha. La lectura del eje vertical, debe hacerse de abajo hacia arriba.
 - ☛ La representación del fenómeno debe variar solo en una dimensión.
 - ☛ En toda gráfica se debe explicar la fuente de donde fueron obtenidos los datos; además, aclarar las escalas, leyendas, notas y convenciones que ayuden a identificar las características presentadas.
 - ☛ Las gráficas deben ser lo bastante sencillas para procurar una idea clara y ser comprensibles sin la ayuda de las descripciones del texto.
 - ☛ Los datos numéricos sobre los que se basa la gráfica deben presentarse en tabla de frecuencia adjunto, si no son incluidos en la misma gráfica.
 - ☛ Las gráficas deben seguir y nunca preceder a la exposición del texto.
-

- ☛ Cuando la gráfica presenta más de una variable, deberá hacerse una muy clara diferenciación por medio de leyendas, notas o signos convencionales.

2.3.1. Diagrama de barras

- Simple: Suministra información sobre una variable únicamente.

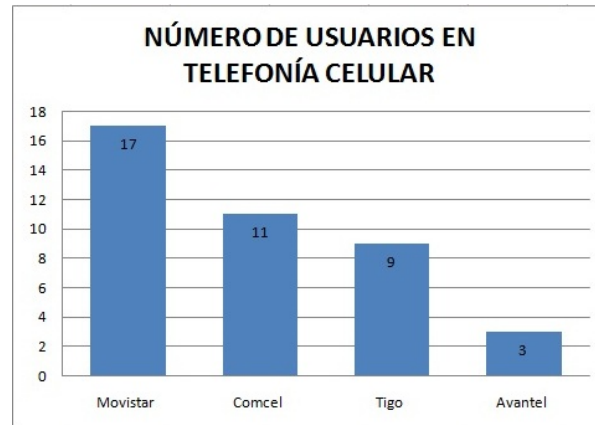
El diagrama de barras es un gráfico que consta de dos ejes: uno de ellos lleva los valores de la variable y el otro representa una frecuencia que puede ser la absoluta o la porcentual. El diagrama de barras se utiliza para variable cualitativa y cuantitativa discreta (tablas de frecuencias sin intervalos) y los rectángulos que lo componen van separados.

Se llama así porque las frecuencias de cada categoría de la distribución se hacen figurar por trazos o columnas de longitud proporcional.

Es una de las graficas más utilizadas por su sencillez.

Las barras para representar las características, por lo general, son construidas en forma vertical, sobre una base horizontal, en la cual se colocan las características por ejemplo el tiempo (años, meses, etc.) y la altura esta dada por los valores o cantidades que toma la variable o el atributo.

La representación se puede hacer utilizando barras horizontales, y se obtiene la misma visualización como cuando las barras son verticales. En ambos casos, la longitud de las barras son proporcionales al número de observaciones o frecuencias, mientras que el ancho de la barra dependerá del gusto de quien lo elabora.

Ejemplo:

El diagrama de barras proporciona información comparativa principalmente y este es su uso principal, por ejemplo de las personas encuestadas 17 son usuarios de Movistar; mientras que 3 son de Avantel.

Ejemplo:

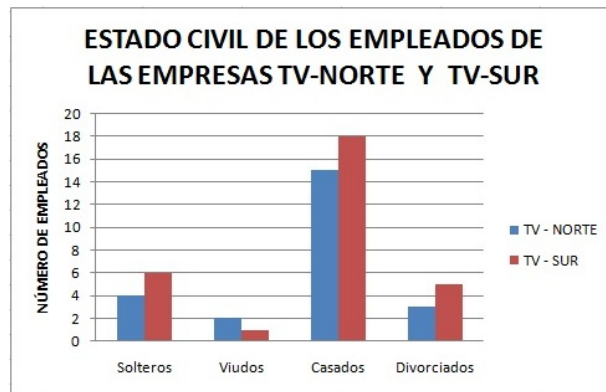
Observando el gráfico, una de las conclusiones que se obtiene es que 20 de las encuestadas cada una tiene un solo hijo.

- Compuesto: Suministra información sobre dos variables y su función principal es comparar.

En ocasiones se tienen datos de dos variables y se quiere representarlos en un mismo diagrama de barras para compararlos, lo más probable es que no haya el mismo número de observaciones en cada una de ellas.

Ejemplo:

Dos empresas estudian el estado civil de sus empleados con el siguiente resultado:



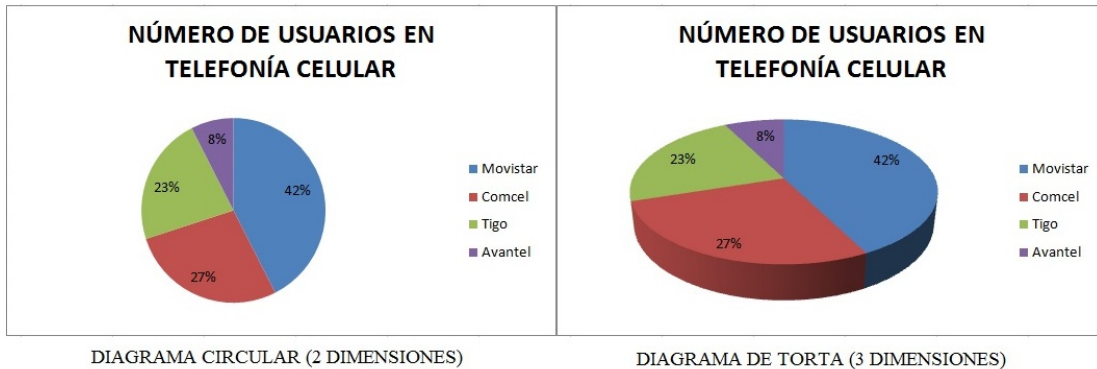
Observando el gráfico, se puede concluir que en la empresa Tv-Sur hay mayor número de empleados casados que en Tv-Norte.

2.3.2. Diagramas circulares

Se utilizan con mucha frecuencia para representar características cualitativas, y sirve para resaltar las diferencias en las proporciones o porcentajes en que está dada la distribución.

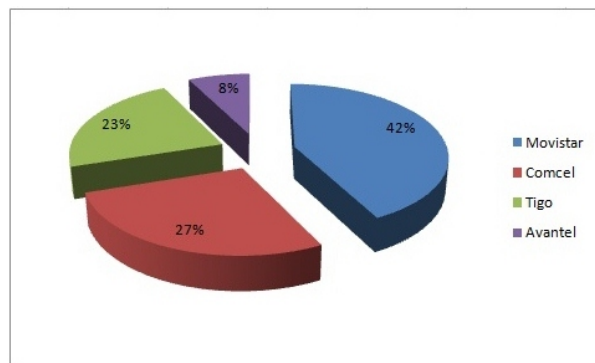
Este tipo de comparación es relativamente efectivo, siempre y cuando los segmentos sean lo suficientemente grandes para permitir comparaciones.

Los datos se representan en un círculo, de modo que el ángulo de cada sector es proporcional a la frecuencia absoluta correspondiente.



Este se usa cuando se trabaja con datos que tienen grandes frecuencias, y los valores de la variable son pocos, la ventaja que tiene este diagrama es que es fácil de hacer y es entendible; la desventaja que posee es que cuando los valores de la variable son muchos es casi imposible o mejor dicho no informa mucho este diagrama y no es productivo, proporciona principalmente información acerca de las frecuencias de los datos de una manera entendible y sencilla.

Cuando el diagrama circular se elabora en tres dimensiones recibe el nombre de torta. Es de gran utilidad cuando se quiere resaltar un determinado valor de la variable, luego se dibuja con el pedazo de la torta separado.

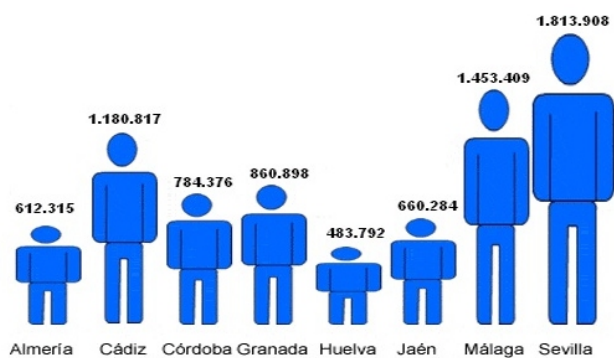


2.3.3. Pictogramas

Expresan con dibujos alusivos al tema de estudio las frecuencias de las modalidades de la variable. Estos gráficos se hacen representando a diferentes escalas un mismo dibujo.

El escalamiento de los dibujos debe ser tal que el área de cada uno de ellos sea proporcional a la frecuencia de la modalidad que representa, el tamaño puede variar y su altura estará dada por la frecuencia absoluta o relativa.

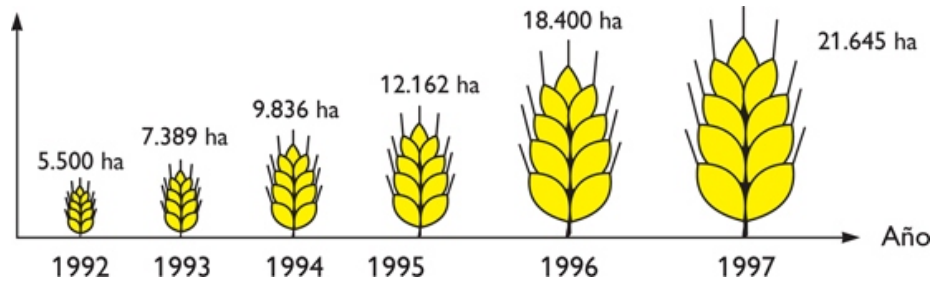
Las figuras empleadas deben explicarse por sí mismas. Se acostumbra que el tamaño sea uniforme, indicándose, aparte de las figuras, el valor de una de ellas; en algunos casos, se acostumbra colocar o señalar el valor total del conjunto, buscando de esta manera la eliminación del cuadro.



Este tipo de gráficos suele usarse en los medios de comunicación, para que sean comprendidos por el público no especializado, sin que sea necesaria una explicación compleja. Así por ejemplo: si se trata de describir la variación sobre el número de estudiantes por colegios en cierta ciudad, se representará cierta cantidad de estudiantes por la figura de un niño o niña.

Los pictogramas bien elaborados, logran obtener efectos positivos y muy atractivos, que llaman la atención. Pero debe desconfiarse de este tipo de gráficos, demasiado recargados, que impiden una mejor visualización de la información presentada.

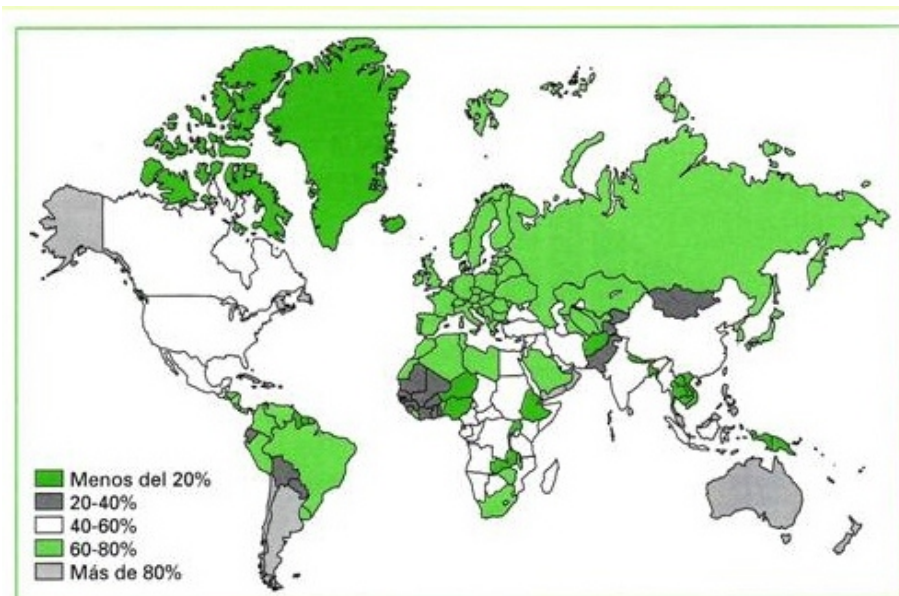
Por ejemplo a continuación se presenta un pictograma en el cual se representa la cantidad de hectáreas cultivadas de trigo durante 6 años.



2.3.4. Cartogramas

Los cartogramas son gráficos realizados sobre mapas, en los que aparecen indicados sobre las distintas zonas cantidades o colores de acuerdo con el carácter que representan.

En el siguiente cartograma se observa la urbanización en el mundo atendiendo a la industrialización:



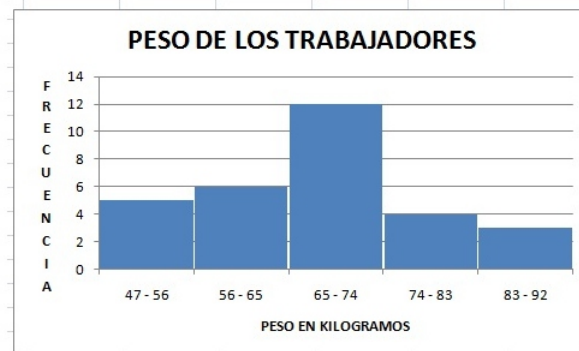
2.3.5. Histograma de frecuencias

Se emplea para ilustrar muestras agrupadas por intervalos. Está formado por un conjunto de rectángulos, cada uno de ellos levantado para cada intervalo, de tal manera que la base será igual a la amplitud C y la altura está dada, ya sea por la frecuencia absoluta o por la

relativa. Sin embargo en la confección de este tipo de gráfica se puede presentar el problema de tener una amplitud no constante, dándonos una gráfica con una imagen engañosa de la distribución que se quiere presentar.

A partir de una tabla de frecuencias se puede construir dos histogramas diferentes, uno para las frecuencias y otro para las frecuencias acumuladas.

L. inf.	L. sup.	y	f	fr	F	Fr	%
47	56	51,5	5	0,17	5	0,17	17
56	65	60,5	6	0,20	11	0,37	20
65	74	69,5	12	0,40	23	0,77	40
74	83	78,5	4	0,13	27	0,90	13
83	92	87,5	3	0,10	30	1,00	10
			30				100

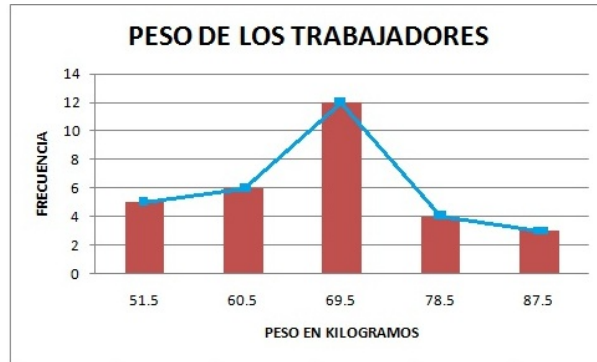


Cuando se elabora un histograma para datos agrupados por intervalos, en el eje horizontal se ubican los límites de los intervalos teniendo en cuenta que el límite superior se escribe solo una vez y los rectángulos que lo comprenden van unidos de límite a límite.

2.3.6. Polígono de frecuencias

El polígono de frecuencias corresponde al diagrama de línea elaborado con los puntos medios de cada intervalo. Con la misma información que fue utilizada para la elaboración del histograma se puede dibujar el polígono de frecuencias.

En primer lugar, se establecen los puntos medios en la parte superior de cada rectángulo, luego se unen dichos puntos mediante una línea recta.



Si la frecuencia que se va a representar es la acumulada, entonces, el polígono recibe el nombre de **ojiva**.



También se puede dibujar el *polígono*, estableciendo las marcas de clases de cada uno de los intervalos, que se colocan en el eje horizontal o abscisa. Para cada valor de la variable corresponderá un valor de la frecuencia, señalándose en el plano por un punto; luego de establecidos todos los puntos, se unen mediante líneas rectas, las que en conjunto forman el polígono.

2.4. Aplicaciones en Excel

2.4.1. Construcción de tablas de frecuencias para variable discreta

Para la construcción de la tabla de frecuencias para una variable discreta se van a utilizar los siguientes datos que corresponden al número de hermanos que tienen 30 estudiantes del colegio “Colombia es Pasión”.

$$\begin{array}{l}
 X_1 = 2 \quad X_7 = 2 \quad X_{13} = 3 \quad X_{19} = 4 \quad X_{25} = 1 \\
 X_2 = 0 \quad X_8 = 1 \quad X_{14} = 3 \quad X_{20} = 4 \quad X_{26} = 2 \\
 X_3 = 2 \quad X_9 = 2 \quad X_{15} = 4 \quad X_{21} = 3 \quad X_{27} = 0 \\
 X_4 = 2 \quad X_{10} = 2 \quad X_{16} = 1 \quad X_{22} = 3 \quad X_{28} = 3 \\
 X_5 = 3 \quad X_{11} = 2 \quad X_{17} = 3 \quad X_{23} = 1 \quad X_{29} = 1 \\
 X_6 = 0 \quad X_{12} = 3 \quad X_{18} = 4 \quad X_{24} = 2 \quad X_{30} = 2
 \end{array}$$

Cada uno de los datos se empiezan a digitar a partir de la casilla **B2** de Excel.

Primero se determina el tamaño de la muestra o población. Para obtener esta información se utiliza la función CONTAR.

La función CONTAR: Cuenta un conjunto de celdas que tengan números en su contenido

Para aplicar la fórmula, primero se crea una casilla en la celda **B10**, que contenga la cantidad de datos (n).

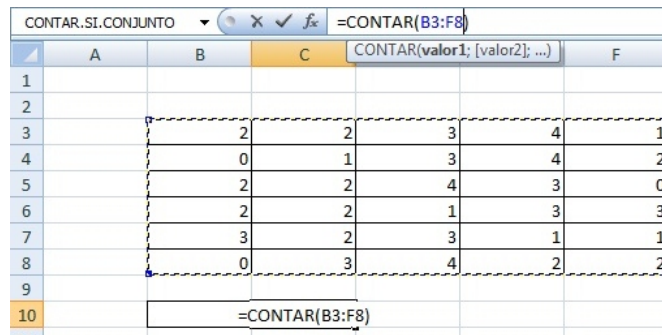
Veamos:

2	2	3	4	1
0	1	3	4	2
2	2	4	3	0
2	2	1	3	3
3	2	3	1	1
0	3	4	2	2
n				

En la celda **C11**, se ingresa la función CONTAR así:

$$= \text{CONTAR}()$$

Luego se ubica en la barra de formulas, con el puntero se dirige a la casilla **B2**, y con click sostenido se desplaza hasta seleccionar todos los datos.



Al pulsar Enter, se tendrá como resultado el conteo de las celdas que tienen números (30 datos en total).

2	2	3	4	1
0	1	3	4	2
2	2	4	3	0
2	2	1	3	3
3	2	3	1	1
0	3	4	2	2
n	30			

Se procede a constuir la tabla de frecuencias para 5 observaciones:

NÚMERO DE HERMANOS	f	fr	F	Fr	%

En la columna número de hermanos, se va a escribir un representante de cada observación y se tiene el siguiente resultado:

NÚMERO DE HERMANOS	f	fr	F	Fr	%
0					
1					
2					
3					
4					
Total					

Ahora hay que determinar las frecuencias absolutas, frecuencias relativas y frecuencia porcentual.

Para hallar la frecuencia absoluta (f). Se empleará una nueva función:

= **CONTAR.SI.CONJUNTO**: Cuenta el número de celdas que cumplen un determinado conjunto de condiciones o criterios.

A partir de la celda **H12**, se activa la función **CONTAR.SI.CONJUNTO** en la cual se le debe suministrar la información de rangocriterios y criterio 1. Para seleccionar rangocriterios, se hace click en la casilla **B2** y con click sostenido se corre hasta la casilla **F7** (lo que se muestra de azul en la figura), luego se suelta el puntero, y en la barra de formulas se digita un punto y coma. Ahora se selecciona criteri 1, se selecciona la casilla **G12** (lo que se muestra con verde en la figura), se suelta el click y se procede a cerrar el paréntesis en la barra de tareas.

NÚMERO DE HERMANOS	f	fr	F	Fr	%
0	=CONTAR.SI.CONJUNTO(B2:F7;G12)				
1	CONTAR.SI.CONJUNTO(rango_criterios1; criterio1; [rango_criterios2; ...])				
2					
3					
4					
Total					

Al dar enter, Excel mostrará la frecuencia para la primera observación:

NÚMERO DE HERMANOS	f	fr	F	Fr	%
0	3				
1					
2					
3					
4					
Total					

Se procede a hacer lo mismo con cada una de los siguientes representates, y se obtiene:

NÚMERO DE HERMANOS	f	fr	F	Fr	%
0	3				
1	5				
2	10				
3	8				
4	4				
Total					

Después en la casilla **H17**, se introduce la formula = *suma()*, y se selecciona desde **H12** hasta **H16**, click y debe aparecer el número total de datos, que en este caso es 30.

- Para calcular las Frecuencias Relativas (fr) se toma cada Frecuencia Absoluta y se divide por el número total de datos (**H17**).

NÚMERO DE HERMANOS	f	fr	F	Fr	%
0	3	=H12/H17			
1	5				
2	10				
3	8				
4	4				
Total	30				

Para poder arrastrar la fórmula se debe fijar primero la celda **H17** (\$H\$17). Después de haberla arrastrado, ahora se puede disminuir los decimales, para poderlo hacer se seleccionan los datos y se dirige a la barra de herramientas y se hace click en disminuir decimales como se muestra en la figura:

The screenshot shows the Excel ribbon with the 'Decrease Decimals' button circled in red. Below it, a table displays the calculated relative frequencies (fr) for each number of siblings.

NÚMERO DE HERMANOS	f	fr	F	Fr	%
0	3	0,10			
1	5	0,17			
2	10	0,33			
3	8	0,27			
4	4	0,13			
Total	30				

La primera Frecuencia Absoluta Acumulada será igual a la primera Frecuencia Absoluta (3). En la celda **J12** se digita el signo igual y luego se pulsa sobre la celda **I12** para trasladar su valor.

La segunda Frecuencia Absoluta Acumulada equivale a la primera Frecuencia Absoluta Acumulada más la segunda Frecuencia Absoluta. En Excel se vería como sigue:

NÚMERO DE HERMANOS	f	fr	F	Fr	%
0	3	0,10	3		
1	5	0,17	=J12+I13		
2	10	0,33			
3	8	0,27			
4	4	0,13			
Total	30				

Se hace click o enter para obtener el resultado, para calcular el resto de Frecuencias, se arrastra la fórmula que está en **J13** hasta la celda **J16**. El resultado final se muestra a continuación:

NÚMERO DE HERMANOS	f	fr	F	Fr	%
0	3	0,10	3		
1	5	0,17	8		
2	10	0,33	18		
3	8	0,27	26		
4	4	0,13	30		
Total	30				

- El cálculo de la Frecuencia Relativa Acumulada (Fr) se hace de forma similar que el cálculo de la Frecuencia Absoluta Acumulada (F).

NÚMERO DE HERMANOS	f	fr	F	Fr	%
0	3	0,10	3	0,10	
1	5	0,17	8	0,27	
2	10	0,33	18	0,60	
3	8	0,27	26	0,87	
4	4	0,13	30	1,00	
Total	30				

- Ahora para calcular el porcentaje, simplemente se multiplica la frecuencia relativa (fr), por 100.

NÚMERO DE HERMANOS	f	fr	F	Fr	%
0	3	0,10	3	0,10	10,00
1	5	0,17	8	0,27	
2	10	0,33	18	0,60	
3	8	0,27	26	0,87	
4	4	0,13	30	1,00	
Total	30				

Para calcular el resto de los porcentajes, simplemente se arrastra la formula hacia abajo, como lo se ha hecho anteriormente. Y se realiza de nuevo el procedimiento de disminuir decimales, para que aparezca lo siguiente:

NÚMERO DE HERMANOS	f	fr	F	Fr	%
0	3	0,10	3	0,10	10,0
1	5	0,17	8	0,27	16,7
2	10	0,33	18	0,60	33,3
3	8	0,27	26	0,87	26,7
4	4	0,13	30	1,00	13,3
Total	30				

En la casilla **L17** se va a calcular la suma total de todos los porcentajes, la cual debe ser igual al 100 %, tal como se observa en la siguiente tabla de frecuencias:

NÚMERO DE HERMANOS	f	fr	F	Fr	%
0	3	0,10	3	0,10	10,0
1	5	0,17	8	0,27	16,7
2	10	0,33	18	0,60	33,3
3	8	0,27	26	0,87	26,7
4	4	0,13	30	1,00	13,3
Total	30				100,0

Finalmente, se le da estética a la tabla y se obtiene la siguiente:

NÚMERO DE HERMANOS	f	fr	F	Fr	%
0	3	0,10	3	0,10	10,0
1	5	0,17	8	0,27	16,7
2	10	0,33	18	0,60	33,3
3	8	0,27	26	0,87	26,7
4	4	0,13	30	1,00	13,3
Total	30				100,0

2.4.2. Construcción de tablas de frecuencias para variable continua (intervalos)

Para la explicación de la creación de una tabla de este tipo en Excel, se van a utilizar los siguientes datos que corresponden al peso en kilogramos de 30 personas.



$$\begin{array}{cccccc} X_1 = 48 & X_7 = 70 & X_{13} = 92 & X_{19} = 85 & X_{25} = 52 & \\ X_2 = 56 & X_8 = 63 & X_{14} = 70 & X_{20} = 68 & X_{26} = 58 & \\ X_3 = 60 & X_9 = 72 & X_{15} = 69 & X_{21} = 82 & X_{27} = 76 & \\ X_4 = 67 & X_{10} = 76 & X_{16} = 61 & X_{22} = 55 & X_{28} = 57 & \\ X_5 = 47 & X_{11} = 74 & X_{17} = 71 & X_{23} = 65 & X_{29} = 72 & \\ X_6 = 70 & X_{12} = 67 & X_{18} = 79 & X_{24} = 88 & X_{30} = 67 & \end{array}$$

Cada uno de los datos se empiezan a digitar a partir de la casilla **B2** de Excel.

Paso 1: Determinar el número de intervalos.

El número de intervalos depende del tamaño de la muestra o población. Para obtener esta información se utiliza la función CONTAR.

La función CONTAR: Cuenta un conjunto de celdas que tengan números en su contenido

Para aplicar la fórmula, primero se crea una tabla resumen a partir de la celda **B10**, que empiece por la cantidad de datos (n) y el número de intervalos

Veamos:

	A	B	C	D	E	F	G
1							
2		48	70	92	85	52	
3		56	63	70	68	58	
4		60	72	69	82	76	
5		67	76	61	55	57	
6		47	74	71	65	72	
7		70	67	79	88	67	
8							
9							
10		n					
11		nc					
12							
13							
14							

En la celda **C11**, se ingresa la función CONTAR así:

$$= \text{CONTAR}()$$

Luego se ubica en la barra de formulas, con el puntero se dirige a la casilla **B2**, y con click sostenido se desplaza hasta seleccionar todos los datos.

	A	B	C	D	E	F	G
1							
2		48	70	92	85	52	
3		56	63	70	68	58	
4		60	72	69	82	76	
5		67	76	61	55	57	
6		47	74	71	65	72	
7		70	67	79	88	67	
8							
9							
10		n					
11		nc	=CONTAR(B2:F7)				
12							

Al pulsar Enter, se tendrá como resultado el conteo de las celdas que tienen números (30 datos en total).

	A	B	C	D	E	F	G
1							
2		48	70	92	85	52	
3		56	63	70	68	58	
4		60	72	69	82	76	
5		67	76	61	55	57	
6		47	74	71	65	72	
7		70	67	79	88	67	
8							
9							
10							
11		n	30				
12		Nc					

Aplicando la fórmula:

$$Nc = \sqrt{n}$$

Se obtendrá el número de intervalos.

Se ubica el cursor en la celda **C12** y se activa esta función: se escribe =raiz(30) y luego enter.

C12		fx =RAIZ(C11)					
	A	B	C	D	E	F	G
1							
2		48	70	92	85	52	
3		56	63	70	68	58	
4		60	72	69	82	76	
5		67	76	61	55	57	
6		47	74	71	65	72	
7		70	67	79	88	67	
8							
9							
10							
11		n	30				
12		Nc	5,47722558				
13							
14							

Al dar Enter se tendrá como resultado el valor 5,477, el cual será igual a 5 indicando que la tabla tendrá 5 intervalos (usando el redondeo).

Paso 2: Determinar el ancho de cada intervalo.

Aumentar la tabla resumen con cuatro nuevas filas: valor mínimo (Xmín.), valor máximo (Xmáx.), rango (R) y ancho del intervalo de clase (A).

Para determinar el valor mínimo y máximo se utilizaran las dos siguientes fórmulas:

= **MIN**: Localiza y muestra el valor mínimo de un conjunto de números.

= **MAX**: Localiza y muestra el valor máximo de un conjunto de números.

En la casilla número1 de la función MIN, se ubicará el rango correspondiente a los datos del ejercicio, luego se procede a hacer lo mismo con la función MAX.

The screenshot shows an Excel spreadsheet with a data table in columns B through F, rows 2 through 7. The formula bar shows the function `=min(B2:F7)`. Below the data table, a summary table is visible:

n	30
Nc	5,477
Xmin	<code>=min(B2:F7)</code>
Xmax	<code>MIN(número1; [número2]; ...)</code>

The screenshot shows the same Excel spreadsheet as above, but with the formula bar showing `=max(B2:F7)`. The summary table below the data table is updated:

n	30
Nc	5,477
Xmin	47
Xmax	<code>=max(B2:F7)</code>

En la casilla **C15** se escribe la fórmula del rango y de igual manera, en la casilla **C16** se escribe la fórmula de la Amplitud de cada intervalo.

	A	B	C	D	E	F	G
1							
2		48	70	92	85	52	
3		56	63	70	68	58	
4		60	72	69	82	76	
5		67	76	61	55	57	
6		47	74	71	65	72	
7		70	67	79	88	67	
8							
9							
10		n	30				
11		nc	5				
12		Xmin	47				
13		Xmax	92				
14		Rango	45				
15		Amplitud	9				

Paso 3: Determinar el nuevo Rango (Rango N).

Se ajusta primero el ancho del intervalo a 9 (para trabajar con valores enteros) con la siguiente función:

= **MULTIPLO.SUPERIOR**: Redondea un número hacia arriba.

Con una nueva fila encabezada por Amplitud C. (indica el ancho ajustado), se activa esta función. En el parámetro número se selecciona la celda en donde se encuentra el ancho sin ajustar (**C16**), y en cifra significativa (equivale al múltiplo al que se desea redondear), el valor de 1.

C18		fx =MULTIPLO.SUPERIOR(C16;1)				
	A	B	C	D	E	F
1						
2		48	70	92	85	52
3		56	63	70	68	58
4		60	72	69	82	76
5		67	76	61	55	57
6		47	74	71	65	72
7		70	67	79	88	67
8						
9						
10						
11		n	30			
12		nc	5			
13		Xmin	47			
14		Xmax	92			
15		Rango	45			
16		Amplitud	9			
17		Rango Nuevo.				
18		Amplitud C.	9			
19						

El nuevo rango resulta de la multiplicación entre la celda **C18** y **C12**:

	A	B	C	D	E	F
1						
2		48	70	92	85	52
3		56	63	70	68	58
4		60	72	69	82	76
5		67	76	61	55	57
6		47	74	71	65	72
7		70	67	79	88	67
8						
9						
10						
11		n	30			
12		nc	5			
13		Xmin	47			
14		Xmax	92			
15		Rango	45			
16		Amplitud	9			
17		Rango Nuevo.	45			
18		Amplitud C.	9			
19						

Paso 4 y paso 5: Determinar los intervalos.

Se procede a constuir la tabla de frecuencias para 5 intervalos de clase:

n	30		L. inf.	L. sup.	y	f	fr	F	Fr	%
nc	5									
Xmin	47									
Xmax	92									
Rango	45									
Amplitud	9									
Rango Nuevo.	45									
Amplitud C.	9									

El límite inferior para el primer intervalo de clase es 47, siendo su límite superior 56 (Xmín más el ancho de clase).

Veamos:

B	C	D	E	F
48	70	92	85	52
56	63	70	68	58
60	72	69	82	76
67	76	61	55	57
47	74	71	65	72
70	67	79	88	67
n	30		L. inf.	L. sup.
nc	5		47	=E12+C18
Xmin	47			
Xmax	92			
Rango	45			
Amplitud	9			
Rango Nuevo.	45			
Amplitud C.	9			

El límite inferior del siguiente intervalo es igual al límite superior del intervalo anterior:

FRECUENCIA		A	B	C	D	E	F
1							
2			48	70	92	85	52
3			56	63	70	68	58
4			60	72	69	82	76
5			67	76	61	55	57
6			47	74	71	65	72
7			70	67	79	88	67
8							
9							
10							
11			n	30		L. inf.	L. sup.
12			nc	5		47	56
13			Xmin	47		=F12	
14			Xmax	92			
15			Rango	45			
16			Amplitud	9			
17			Rango Nuevo.	45			
18			Amplitud C.	9			

Con el cursor se arrastra esta fórmula hasta el último intervalo de clase (**E16**):

	A	B	C	D	E	F
1						
2		48	70	92	85	52
3		56	63	70	68	58
4		60	72	69	82	76
5		67	76	61	55	57
6		47	74	71	65	72
7		70	67	79	88	67
8						
9						
10						
11		n	30		L. inf.	L. sup.
12		nc	5		47	56
13		Xmin	47		56	
14		Xmax	92		0	
15		Rango	45		0	
16		Amplitud	9		0	
17		Rango Nuevo.	45			
18		Amplitud C.	9			

Para calcular los límites superior bastará con arrastrar la fórmula que está en la celda **F17**, fijando de antemano, la celda **C18** que hace referencia al tamaño de la clase:

FRECUENCIA						
	A	B	C	D	E	F
1						
2		48	70	92	85	52
3		56	63	70	68	58
4		60	72	69	82	76
5		67	76	61	55	57
6		47	74	71	65	72
7		70	67	79	88	67
8						
9						
10						
11		n	30		L. inf.	L. sup.
12		nc	5		47	=E12+\$C\$18
13		Xmin	47		56	
14		Xmax	92		0	
15		Rango	45		0	
16		Amplitud	9		0	
17		Rango Nuevo.	45			
18		Amplitud C.	9			

El resultado de los intervalos es el siguiente:

L. inf.	L. sup.	y	f	fr	F	Fr	%
47	56						
56	65						
65	74						
74	83						
83	92						

Paso 6: Determinar las frecuencias absolutas, frecuencias relativas y marcas de clases.

Para hallar la marca de clase, se suma el límite inferior con el límite superior y el resultado se divide entre 2.

L. inf.	L. sup.	y	f	fr	F	Fr	%
47		$=(E12+F12)/2$					
56	65						
65	74						
74	83						
83	92						

Después se arrastra hacia abajo la formula, como se ve en la siguiente figura:

L. inf.	L. sup.	y	f	fr	F	Fr	%
47	56	51,5					
56	65	60,5					
65	74	69,5					
74	83	78,5					
83	92	87,5					

Ahora se halla la frecuencia absoluta (f). Para trabajar con intervalos se empleará una nueva función:

= **FRECUENCIA**: Muestra el número de veces que se repite un número dentro de un rango de celdas.

A partir de la celda **H12**, se activa la función FRECUENCIA en la cual se le debe suministrar la información de datos y grupos. Para seleccionar los datos, se hace click en la casilla **B2** y con click sostenido se corre hasta la casilla **F7**, luego se suelta el puntero, y en la barra de formulas se digita un punto y coma. Ahora se seleccionan los grupos, de igual manera: se selecciona desde la casilla **F12** y con click sostenido se dirige hasta la casilla **G12**, se suelta el click y se procede a cerrar el paréntesis en la barra de tareas.

L. inf.	L. sup.	y	f	fr	F	Fr	%
47	56	$=FRECUENCIA(B2:F7;F12)$					
56	65	60,5					
65	74	69,5					
74	83	78,5					
83	92	87,5					

Al dar enter, Excel mostrará la frecuencia para el primer intervalo de clase:

La primera Frecuencia Absoluta Acumulada será igual a la primera Frecuencia Absoluta (5). En la celda **J12** se digita el signo igual y luego se pulsa sobre la celda **I12** para trasladar su valor.

La Frecuencia Absoluta Acumulada para el segundo intervalo equivale a la Frecuencia Absoluta Acumulada del primer intervalo más la Frecuencia Absoluta del segundo intervalo. En Excel se vería como sigue:

L. inf.	L. sup.	y	f	fr	F	Fr	%
47	56	51,5	5	0,17	5		
56	65	60,5	6	0,20	11		
65	74	69,5	12	0,40			
74	83	78,5	4	0,13			
83	92	87,5	3	0,10			
			30				

Se hace click un enter para obtener el resultado, para calcular el resto de Frecuencias, se arrastra la fórmula que está en **J13** hasta la celda **J16**. El resultado final se muestra a continuación:

L. inf.	L. sup.	y	f	fr	F	Fr	%
47	56	51,5	5	0,17	5		
56	65	60,5	6	0,20	11		
65	74	69,5	12	0,40	23		
74	83	78,5	4	0,13	27		
83	92	87,5	3	0,10	30		
			30				

- El cálculo de la Frecuencia Relativa Acumulada (Fr) se hace de forma similar que el cálculo de la Frecuencia Absoluta Acumulada (F).

L. inf.	L. sup.	y	f	fr	F	Fr	%
47	56	51,5	5	0,17	5	0,17	
56	65	60,5	6	0,20	11		
65	74	69,5	12	0,40	23		
74	83	78,5	4	0,13	27		
83	92	87,5	3	0,10	30		
			30				

L. inf.	L. sup.	y	f	fr	F	Fr	%
47	56	51,5	5	0,17	5	0,17	
56	65	60,5	6	0,20	11	0,37	
65	74	69,5	12	0,40	23	0,77	
74	83	78,5	4	0,13	27	0,90	
83	92	87,5	3	0,10	30	1,00	
			30				

- Ahora para calcular el porcentaje, simplemente se multiplica la frecuencia relativa (fr), por 100.

L. inf.	L. sup.	y	f	fr	F	Fr	%
47	56	51,5	5	0,17	5	0,17	16,6666667
56	65	60,5	6	0,20	11	0,37	
65	74	69,5	12	0,40	23	0,77	
74	83	78,5	4	0,13	27	0,90	
83	92	87,5	3	0,10	30	1,00	
			30				

Para calcular el resto de los porcentajes, simplemente se arrastra la formula hacia abajo, como lo se ha hecho anteriormente. Y se realiza de nuevo el procedimiento de disminuir decimales, para que aparezca lo siguiente:

L. inf.	L. sup.	y	f	fr	F	Fr	%
47	56	51,5	5	0,17	5	0,17	17
56	65	60,5	6	0,20	11	0,37	20
65	74	69,5	12	0,40	23	0,77	40
74	83	78,5	4	0,13	27	0,90	13
83	92	87,5	3	0,10	30	1,00	10
			30				

En la casilla **L17** se va a calcular la suma total de todos los porcentajes, la cual debe ser igual al 100 %, tal como se observa en la siguiente tabla de frecuencias:

L. inf.	L. sup.	y	f	fr	F	Fr	%
47	56	51,5	5	0,17	5	0,17	17
56	65	60,5	6	0,20	11	0,37	20
65	74	69,5	12	0,40	23	0,77	40
74	83	78,5	4	0,13	27	0,90	13
83	92	87,5	3	0,10	30	1,00	10
			30				100

Finalmente, se le da estética a la tabla y se obtiene la siguiente:

L. inf.	L. sup.	y	f	fr	F	Fr	%
47	56	51,5	5	0,17	5	0,17	17
56	65	60,5	6	0,20	11	0,37	20
65	74	69,5	12	0,40	23	0,77	40
74	83	78,5	4	0,13	27	0,90	13
83	92	87,5	3	0,10	30	1,00	10
			30				100

2.4.3. Construcción de gráficos en Excel

Un gráfico es una representación de los datos de una hoja de cálculo a través de figuras o líneas que permiten un análisis e interpretación más claros de los mismos.



Excel tiene una herramienta que permite la construcción de gráficos simples y complejos, brindando claridad en el momento de analizar la información.

En general la representación grafica de los datos hace que estos se vean más interesantes atractivos y fáciles de leer.

Excel provee varios tipos de gráficos con lo cual el usuario puede elegir el más adecuado para cada situación.

PROCESO DE CREACIÓN DE UN GRÁFICO.

Para poder crear cualquier tipo de grafico en Excel, primero se debe tener una tabla de frecuencias como las que se mostraron anteriormente, en la cual deben estar escritos los datos que se quieren representar gráficamente.

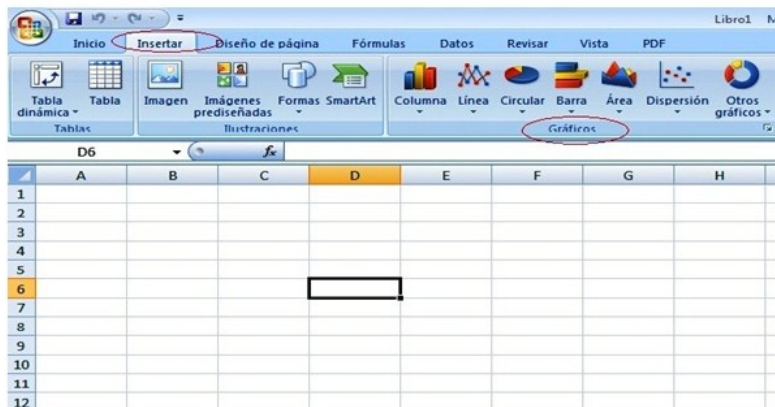
Veamos un ejemplo.

	A	B	C	D	E	F	G
1							
2			¿Qué tipo de Lectura le gusta practicar en su tiempo libre?				
3			Tipo de lectura	f	fr	%	
4			Periodico	13	0,520	52	
5			Libros	0	0,000	0	
6			Revistas	6	0,240	24	
7			No contesto	6	0,240	24	
8				25			

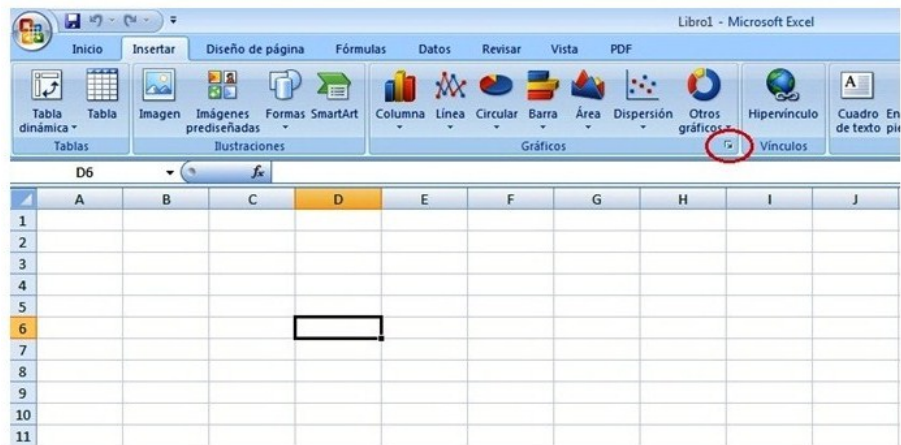
Luego de tener la información en la tabla, se seleccionan los datos. Se recomienda tener seleccionados los datos que se quieran graficar antes de crear el gráfico. De lo contrario el gráfico se mostrará en blanco y se tendran que ingresar las series de datos después.

	A	B	C	D	E	F	G
1							
2			¿Qué tipo de Lectura le gusta practicar en su tiempo libre?				
3			Tipo de lectura	f	fr	%	
4			Periodico	13	0,520	52	
5			Libros	0	0,000	0	
6			Revistas	6	0,240	24	
7			No contesto	6	0,240	24	
8				25			

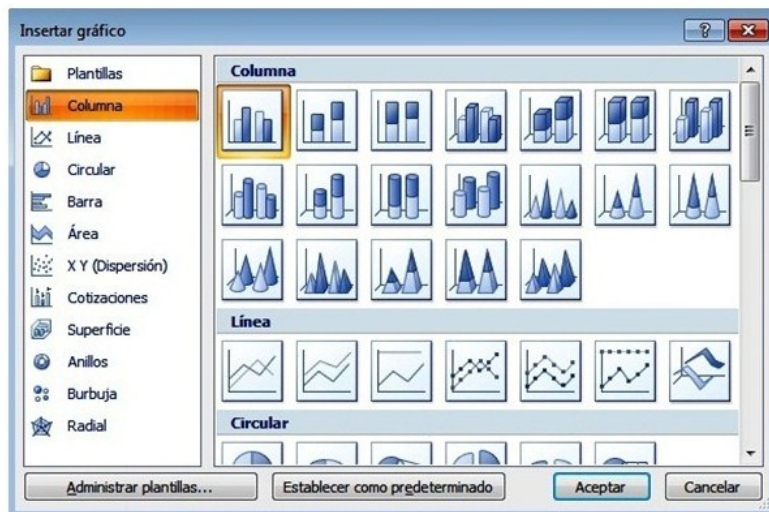
Se hace click es la pestaña insertar, donde se muestran los diferentes gráficos que dispone Excel.



Luego, se hace click en la pestaña de la opción gráfico (tal como se muestra en la siguiente imagen)

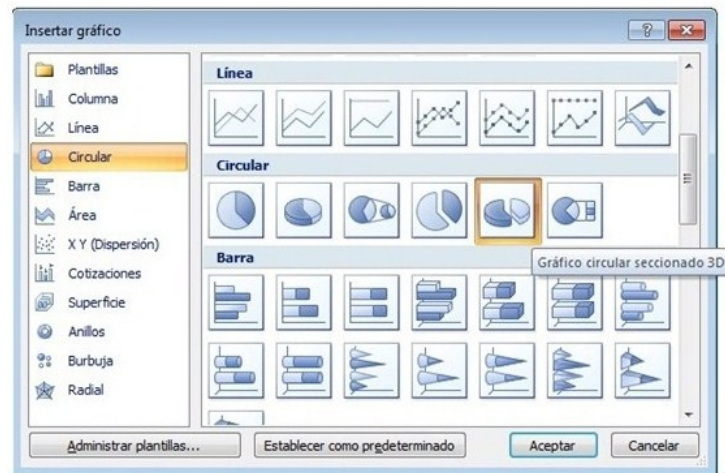


Al darle click, en diagrama de columna aparece una ventana donde se muestran todos los gráficos de columna de los que dispone Excel.

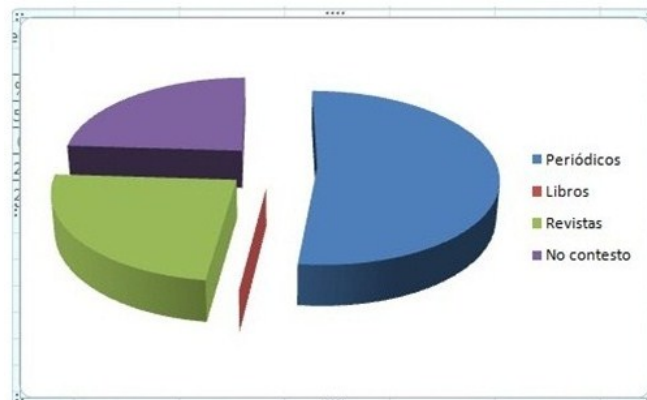


En la parte izquierda de la ventana se muestran las diferentes opciones de los gráficos, donde solamente se puede elegir uno. En la parte derecha se muestra cada una de las variantes del tipo de gráfico que se ha elegido en la parte izquierda. Después de elegir el tipo de gráfico se hace click en aceptar y automáticamente aparece en la hoja de cálculo el grafico final.

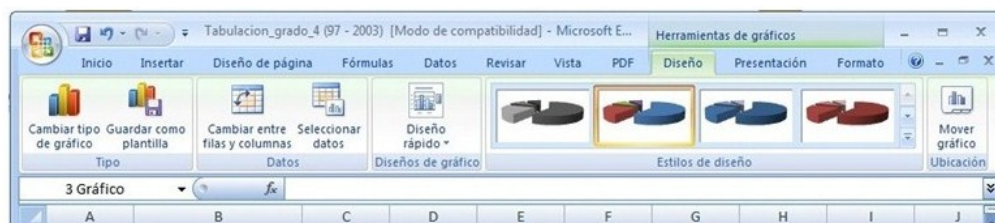
Para esta aplicación se va a diseñar un grafico de sectores o circular.



Después de dar click en aceptar aparece lo siguiente.



En la parte superior de la pantalla de Excel, aparecen opciones para mejorar el gráfico, todo en la pestaña diseño.



Se cuenta con las siguientes:

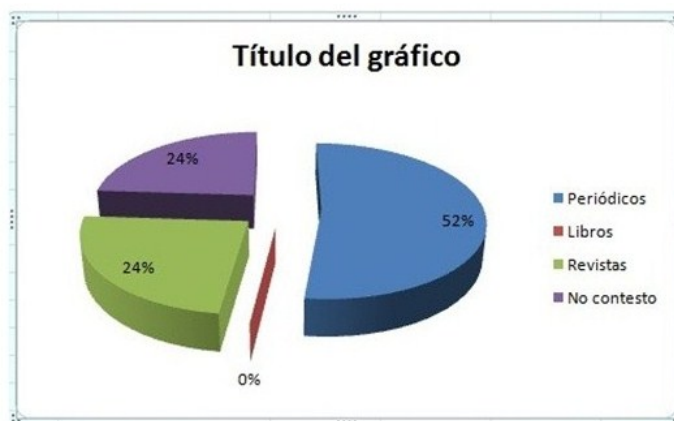
- * Cambiar tipo de gráfico: Esta opción permite cambiar el tipo de gráfico sin tener que eliminar el que se tiene, ni volver a seleccionar los datos de origen del gráfico.

- * Datos: Con esta opción se pueden cambiar los datos con los cuales se puede realizar un nuevo gráfico.
- * Diseño de grafico: Con esta opción se cambia el diseño general del gráfico, se puede incluir el titulo, colocarle etiquetas (porcentaje o cantidad de datos).
- * Estilos de diseños: Con esta opción se puede cambiar el estilo del gráfico, es decir, cambiarle el color para adecuarlo a las necesidades, Excel presenta 48 estilos diferentes.
- * Mover grafico: Con esta opción se puede mover el gráfico a una hoja nueva.

Para este caso se va a agregar el titulo al gráfico.

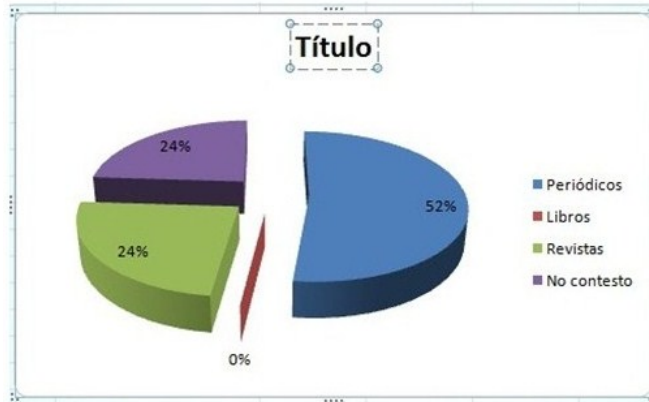
Veamos como se procede:

Se selecciona el Diseño 6. Con lo cual debe aparecer lo siguiente, (para poder hacer este cambio se debe tener seleccionado el gráfico, lo cual se hace simplemente dando click izquierdo sobre él).



Se puede observar que, el formato de este gráfico es mucho más completo e informativo que el anterior, con este nuevo formato le puede incluir un titulo adecuado al gráfico y sobre cada uno de los sectores se muestra el porcentaje al cual equivale cada una de las porciones.

Para escribir el titulo, se hace click izquierdo sobre "Título del grafico", con lo cual se activa un rectángulo en el cual se borra el titulo que tiene.



Ahora se escribe el título que esté acorde con la tabla de frecuencias.



El procedimiento que se ha realizado para mejorar el gráfico se puede hacer con cada uno de los tipos de gráficos existentes.

Creación de un histograma en Excel 2007

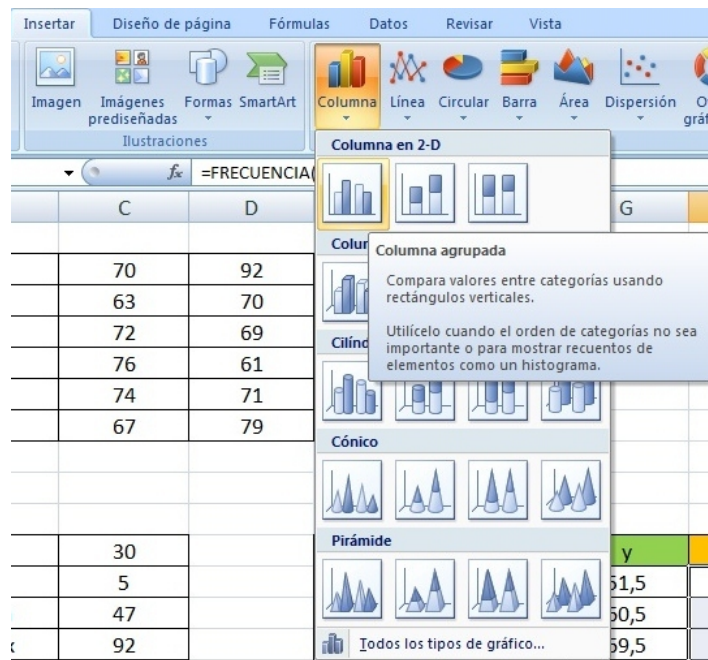
Para la creación del histograma se van a utilizar los datos de la tabla de frecuencias que muestra el peso en Kilogramos de 30 trabajadores.

L. inf.	L. sup.	y	f	fr	F	Fr	%
47	56	51,5	5	0,17	5	0,17	17
56	65	60,5	6	0,20	11	0,37	20
65	74	69,5	12	0,40	23	0,77	40
74	83	78,5	4	0,13	27	0,90	13
83	92	87,5	3	0,10	30	1,00	10
			30				100

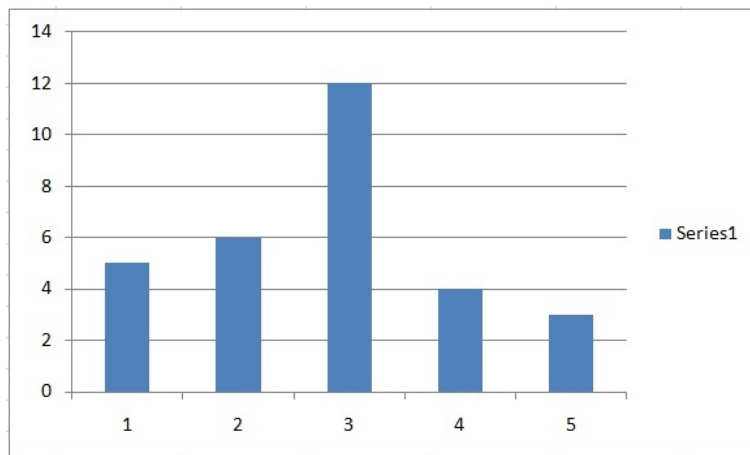
Primero, se seleccionan los datos de las frecuencias relativas

L. inf.	L. sup.	y	f	fr	F	Fr	%
47	56	51,5	5	0,17	5	0,17	17
56	65	60,5	6	0,20	11	0,37	20
65	74	69,5	12	0,40	23	0,77	40
74	83	78,5	4	0,13	27	0,90	13
83	92	87,5	3	0,10	30	1,00	100
			30				

Se hace click en insertar, luego en columna y por último en columna agrupada



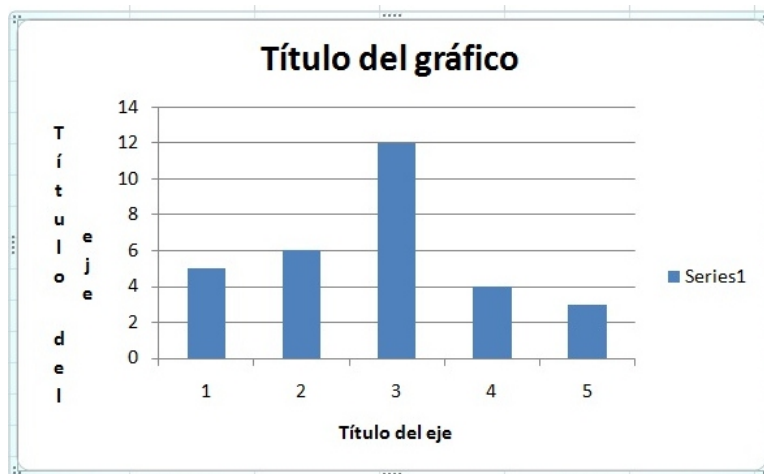
Después de dar click en aceptar se obtiene el siguiente diagrama.



Dando click izquierdo sobre el grafico se dirige a la pestaña de presentación y se da click e inmediatamente aparecen las siguientes opciones



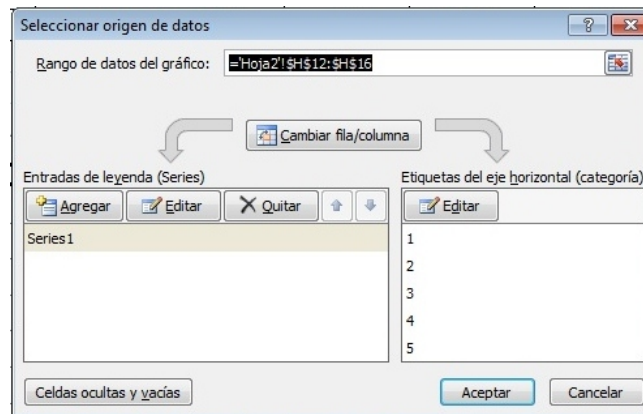
Luego click en título del grafico para incluir el título, después click en Rótulos del eje para poder escribir las características de cada uno de los ejes, en título de eje horizontal primario se da click en título bajo el eje; y en título del eje vertical primario se selecciona título vertical (recuerde que el gráfico original debe estar seleccionado), con lo cual aparece el diagrama con mas información:



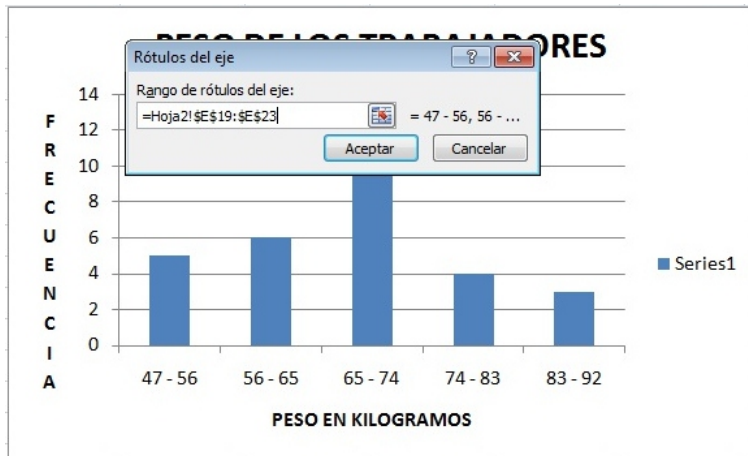
Ahora en Título del gráfico se va a escribir “PESO DE LOS TRABAJADORES”, en el eje horizontal se escribe “PESO EN KILOGRAMOS” y en título del eje vertical se puede escribir “FRECUENCIA”, y queda:



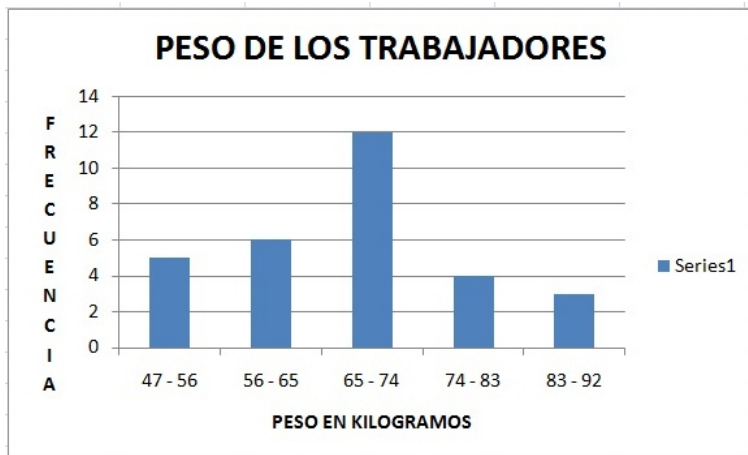
Después se le van a cambiar las categorías del eje horizontal y se le van a poner los límites de cada uno de los intervalos, para poderlo hacer debemos seleccionar el gráfico y luego se dirige hacia la pestaña diseño y click en seleccionar datos, con lo cual aparece una nueva ventana.



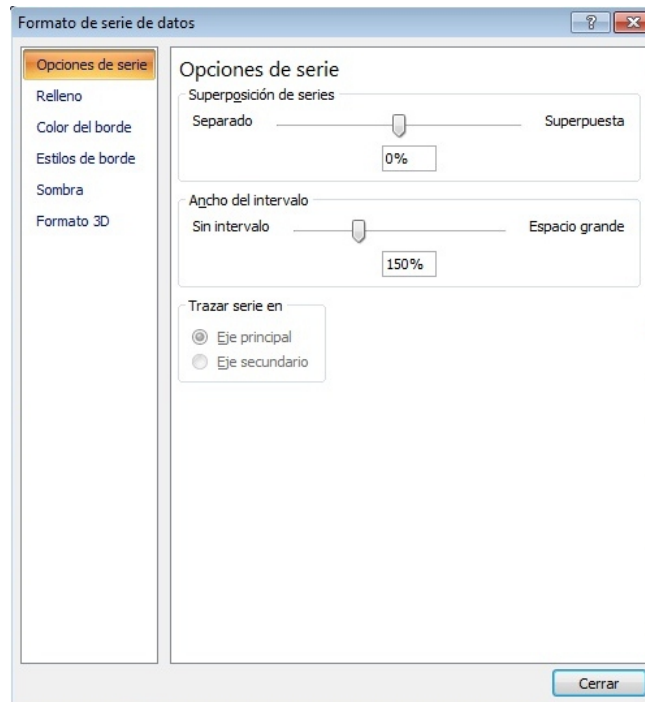
Donde dice etiquetas del eje horizontal se da click en editar y aparece una nueva ventana que dice rango de rótulo del eje, y ahora se seleccionan los límites del intervalo.



Si se observa muy detenidamente el gráfico se puede ver que las categorías del eje horizontal han cambiado y ahora aparecen los límites de los intervalos, se da click en aceptar, nuevamente en aceptar y aparece lo siguiente:



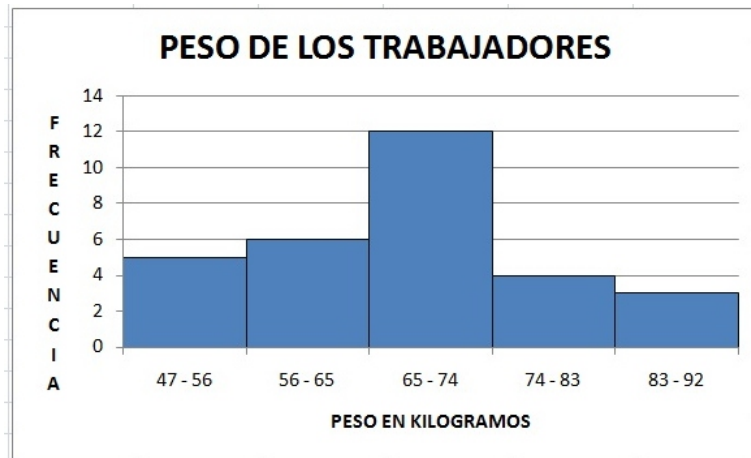
Ahora se van unir la barras entre si para poder que sea un histograma, para esto se da un click sobre una de las barras del gráfico de tal forma que ellas queden seleccionadas, luego sobre ellas se da click derecho aparece una nueva ventana y de ella selecciona dar formato a serie de datos con lo cual aparece una nueva ventana:



Donde dice ancho del intervalo se pone en cero y luego se da click en cerrar.



Ahora se elimina la palabra “Series 1”, se le resaltan los bordes de las barras utilizando nuevamente la opción dar formato a serie de datos pero ahora se da click en la pestaña color del borde, se selecciona línea sólida y se elige el color negro, finalmente se da click en cerrar y aparece el histograma terminado.



Creación de un polígono de frecuencias en Excel 2007

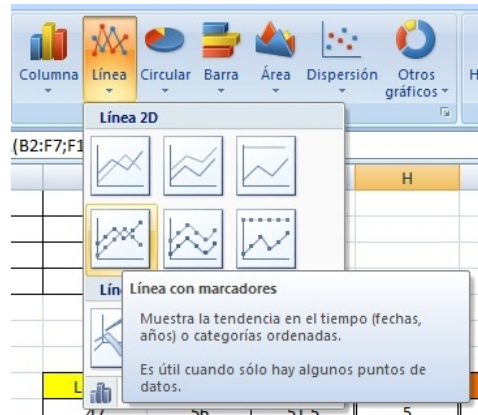
Para la creación del polígono de frecuencias se van a utilizar los mismos datos que se utilizaron para la creación del histograma.

L. inf.	L. sup.	y	f	fr	F	Fr	%
47	56	51,5	5	0,17	5	0,17	17
56	65	60,5	6	0,20	11	0,37	20
65	74	69,5	12	0,40	23	0,77	40
74	83	78,5	4	0,13	27	0,90	13
83	92	87,5	3	0,10	30	1,00	10
			30				100

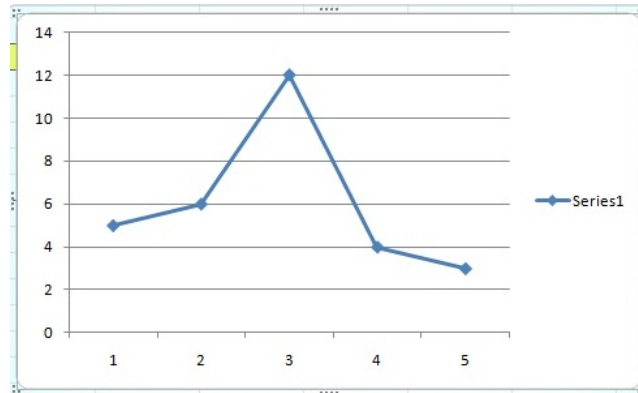
Primero, se seleccionan los datos de las frecuencias absolutas.

L. inf.	L. sup.	y	f	fr	F	Fr	%
47	56	51,5	5	0,17	5	0,17	17
56	65	60,5	6	0,20	11	0,37	20
65	74	69,5	12	0,40	23	0,77	40
74	83	78,5	4	0,13	27	0,90	13
83	92	87,5	3	0,10	30	1,00	10
			30				100

Luego, click en insertar, luego en línea y por último en línea con marcadores.



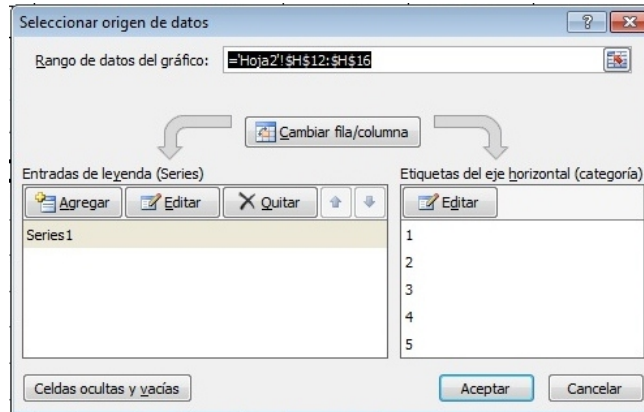
Después de dar click aparece el polígono.



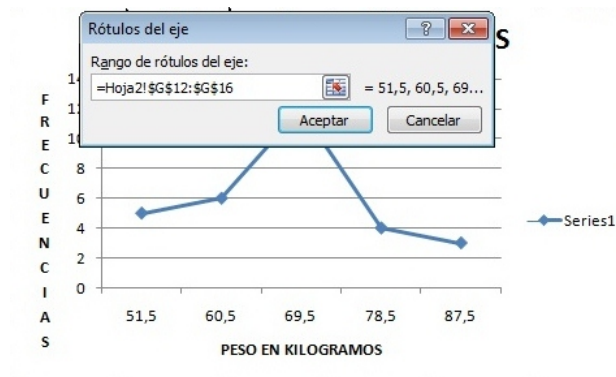
Ahora para colocarle los títulos de cada uno de los ejes y el título general del gráfico se siguen los mismos pasos que con el histograma de frecuencia (estos pasos se pueden hacer para cualquier tipo de gráfico) por tal motivo no se van a explicar. Al realizar esos pasos y después de ponerle los mismos nombres aparece el siguiente polígono de frecuencias



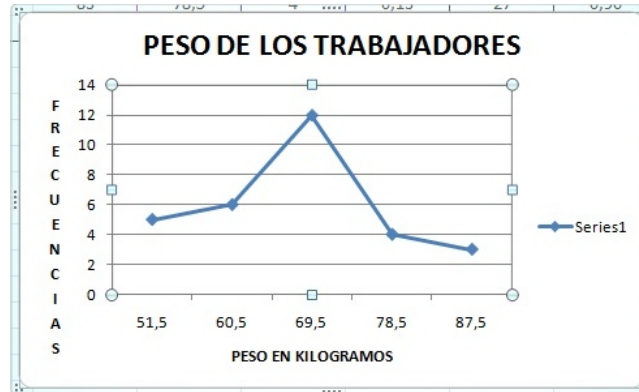
Después se le van a cambiar las categorías del eje horizontal y se le van a poner las marcas de clase de cada uno de los intervalos, para poderlo hacer se debe seleccionar el gráfico y luego se dirige hacia la pestaña diseño y click en seleccionar datos, con lo cual aparece una nueva ventana.



Donde dice etiquetas del eje horizontal se da click en editar y aparece una nueva ventana que dice rango de rótulo del eje, y ahora se seleccionan las marcas de clase de los intervalos.



Si se observa muy detenidamente el gráfico se puede ver que las categorías del eje horizontal han cambiado y ahora aparecen las marcas de clase de cada intervalo, se da click en aceptar, nuevamente en aceptar y se obtiene.



Ahora se elimina la palabra “Series 1”, se cambia el formato al polígono de frecuencias y se selecciona el Diseño 7, en la opción diseños de gráfico, en estilos de diseño se selecciona el Estilo 4 (todo esto se debe hacer teniendo seleccionado el gráfico) y aparece el polígono terminado. (Es posible que al seleccionar el nuevo diseño se alteren los títulos que hemos escrito, si esto ocurre simplemente se vuelven a realizar los pasos anteriores).



También se puede seleccionar el Diseño 9, y si se pierden los títulos se le vuelven a insertar como se hizo anteriormente y queda otro polígono de frecuencias.



2.5. Talleres de aplicaciones

1. En una liga de Fútbol se realizó un estudio sobre las anotaciones hechas en una temporada. Los resultados se muestran en la siguiente tabla:

Número de anotaciones	f	F	f_r	F_r	%
0			0,0921		
1					18,68
2	85				
3			0,2316		
4					14,47
5			0,0632		
6					3,42
7					
Total					

- a. Completar la tabla teniendo en cuenta los datos suministrados.
 - b. Representar los datos de la distribución de frecuencias en un diagrama de barras.
 - c. Puede usted decir ¿cuántas anotaciones constituyen el 80 % de los datos?
2. Determinar la población y la muestra en cada situación.
- a. Un fabricante de medicamentos quiere saber la cantidad de personas cuya hipertensión puede ser controlada con un nuevo producto fabricado por la compañía.
 - b. En un estudio a un grupo de 13000 individuos hipertensos, se encontró que el 80 % de ellos controló su presión con el nuevo medicamento.
 - c. El dueño de una papelería quiere saber si es lucrativo implementar el servicio de fotocopidora. Para ello preguntó a 68 clientes que ingresaron a la papelería si usarían el servicio de fotocopidora.
-

3. Se realiza un estudio para registrar en 18 días los niveles de agua de un río al pasar por la zona sur de la ciudad. Los resultados en metros cúbicos fueron los siguientes:

25 25 26 27 32 33 34 28 25
26 22 28 28 36 34 29 27 26

Se pide:

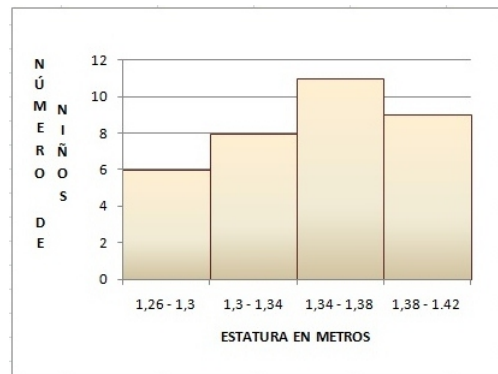
- Determinar la variable del estudio, el tipo de variable y su dominio.
 - Construir la distribución de frecuencias.
 - Elaborar el diagrama de barras para la frecuencia absoluta acumulada.
 - Hallar el porcentaje de días en el que el nivel del agua sobrepasó los $32m^3$.
 - Hallar el porcentaje de días en el que el nivel del agua estuvo entre 25 y $34m^3$.
4. En el aeropuerto, se encuesta a 30 pasajeros internacionales para saber el número de dólares que ingresan al país en cada viaje, las respuestas obtenidas fueron:

30 20 25 10 15 25
35 20 15 10 25 30
20 15 25 35 20 10
12 13 45 25 15 18
23 13 12 15 20 15

Se pide:

- Elaborar una distribución de frecuencias con cinco intervalos.
 - Representar los datos en un histograma porcentual, en un polígono de frecuencias relativas y en una ojiva.
 - Explicar cual es la diferencia entre las dos frecuencias más altas.
-

- d. Determinar el valor con mayor frecuencia relativa.
5. Determinar el tipo de variable en cada situación:
- Ventas mensuales en un almacén
 - Formas de pago de un carro
 - Lugar de vacaciones preferido
 - Marca de una motocicleta
 - Género de musica favorito
6. En un jardín estudiantil se elabora un registro sobre las estaturas de los niños. Los resultados se pueden apreciar en el siguiente grafico:



- Determinar: población, tamaño de la población, muestra, variable y tipo de variable.
 - Elaborar una distribución de frecuencias con intervalos.
7. Los siguientes datos representan el tiempo, en minutos, que esperan 15 pacientes para ser atendidos en la sala de urgencias de un hospital:
- 15, 20, 30, 15, 25, 5, 50, 18, 30, 48, 27, 40, 35, 38, 18

- Realizar la tabla de frecuencias.
-

- Representar los datos en un diagrama circular.
 - Se puede saber: ¿Cuántas personas fueron atendidas entre 10 y 20 minutos?.
 - Determinar si la cantidad de personas que fueron atendidas entre 20 y 30 minutos fue mayor que la cantidad de personas atendidas entre 30 y 40 minutos.
8. Un nadador de 200 metros registra el tiempo de sus últimos 14 entrenamientos, los resultados en segundos fueron los siguientes:

125, 120, 130, 135, 125, 115, 116, 122, 117, 115, 132, 121, 133, 119

Realizar:

- Una tabla de frecuencias con intervalos
 - Una tabla de frecuencias sin intervalos
 - Representar los datos en un diagrama de barras
 - Realizar un diagrama circular
9. Un equipo de futbol realizó pruebas físicas a sus integrantes y registró las pulsaciones después de realizar las pruebas físicas, obteniendo los siguientes resultados:

<i>Intervalo</i>	<i>f</i>
[65 – 70)	12
[70 – 75)	17
[75 – 80)	12
[80 – 85)	29
[85 – 90)	35
[90 – 95)	1

Se pide determinar

- a. ¿qué porcentaje de los deportistas tuvo menos de 85 pulsaciones?
-

MEDIDAS ESTADÍSTICAS

Las tablas de frecuencias y las representaciones gráficas permiten tener una idea global de la distribución de frecuencias, dentro de la población considerada. Pero esta idea global es más bien cualitativa pero también interesa un *resumen cuantitativo* del fenómeno estadístico, de tal manera que sirva para estudios posteriores y para comparación con otras distribuciones.

En numerosas ocasiones, en lugar de trabajar con todos los datos, es preferible disponer de una o más medidas descriptivas que resumen los datos de forma cuantitativa. Estas medidas se llaman Medidas Estadísticas (a veces se utiliza el término estadístico cuando se hace referencia a una muestra). La Estadística Descriptiva es la rama de la Estadística que se encarga del recuento, ordenación y clasificación de los datos y es la que define y calcula los parámetros estadísticos que caracterizan la distribución. Se trata, pues, de resumir en unos cuantos valores la información más importante y útil contenida en un conjunto de datos, haciendo más fácil la comprensión y la comparación de las diversas características de las poblaciones o muestras.

3.1. Medidas de Tendencia central para datos no agrupados

Los datos de un estudio estadístico al igual que las personas y los objetos, en su proceso de organización, necesitan ser distinguidos por sus características. Sin embargo, como no es

posible representarlos en su totalidad, es preciso seleccionar uno que distinga todo el grupo y que en sí mismo reúna esas características.

Para identificar el representante de un conjunto de datos, existen tres medidas muy representativas como: la *media aritmética*, la *mediana* y la *moda*.

Las medidas de tendencia central o de centralización, pretenden encontrar un punto centro o punto de acumulación de los datos.

3.1.1. Media aritmética

Analizando lo que le sucede a María con las notas del curso de Estadística:

Durante el transcurso del semestre en la universidad María obtuvo las siguientes notas: 3,5; 4,0; 3,6; 4,8; 3,0. ¿Cuál es la nota promedio de María?

Para saberlo, se debe encontrar una nota que sea representativa a las 5 notas anteriores. Es decir:

$$\frac{3,5 + 4,0 + 3,6 + 4,8 + 3,0}{5} = \frac{18,9}{5} = 3,78$$

La nota definitiva es 3,78 y recibe el nombre de **promedio** o **media aritmética**. Veamos ordenando 3,0; 3,5; 3,6; 4,0; y 4,8, se nota que 3,78 esta ubicado hacia el centro.

La **media aritmética** (\bar{X}) es un dato que no necesariamente está en el conjunto de datos y representa la característica predominante del grupo. Y se puede considerar como el dato que mejor representa el conjunto, debido a que se trata de ubicar hacia el centro de ellos.

En un conjunto n de observaciones $x_1, x_2, x_3, \dots, x_n$, la media está dada por:

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

donde:

\sum (sigma) es la letra griega que se utiliza para representar sumatoria.

x_i representa cada uno de los datos.

Ejemplo:

El peso en kilogramos de los profesores del colegio “ Colombia eres bella” son los siguientes:

59 54 62 49 65 68 85 74 67 63 42

¿Cuál es el peso promedio de los profesores?

Solución

Para calcular el peso promedio de los profesores del colegio, se suman todos los pesos y se dividen entre el número total de profesores.

$$\bar{X} = \frac{59 + 54 + 62 + 49 + 65 + 68 + 85 + 74 + 67 + 63 + 42}{11} = \frac{688}{11} = 62,55 \text{ Kilogramos}$$

Entonces, el peso promedio de los profesores es 62,55 Kilogramos.

NOTA: La media aritmética es muy sensible a los valores extremos y por ende deja de ser un valor representativo.

Ejemplo:

Las siguientes son las edades de 15 trabajadores de una empresa de lacteos de la ciudad de Neiva.

23 25 24 25 21 20 22 23 23 22 23 22 21 50 55

1. Calcular la edad promedio de los trabajadores.
 2. Calcular la edad promedio de los trabajadores sin tener en cuenta las dos últimas edades.
-

Solución

1. Para hallar la edad promedio se deben sumar las edades y dividir las entre el número de datos, veamos:

$$\bar{X} = \frac{23 + 25 + 24 + 25 + \dots + 22 + 21 + 50 + 55}{15} = \frac{399}{15} = 26,6 \text{ años}$$

2. Ahora se halla el promedio sin tener en cuenta 50 y 55.

$$\bar{X} = \frac{23 + 25 + 24 + 25 + \dots + 23 + 22 + 21}{13} = \frac{294}{13} = 22,6 \text{ años}$$

Como se observa la edad promedio de todas edades se aleja mucho de los valores de las edades, pero cuando se sacan los dos valores extremos el promedio de las edades se acerca más a la realidad.

Por problemas como el anterior con valores extremos, es que se recurre a una nueva medida de tendencia central.

3.1.2. Mediana

El peso en kilogramos de los integrantes del equipo de Fútbol de salón del colegio “Colombia te quiero” son los siguientes:

55 58 60 62 63 68 72 74 74 85 87

Si se observa muy detenidamente los datos, se puede percibir que se encuentran ordenados de manera ascendente y también que hay un número impar de datos 11, y que en toda la mitad se encuentra la posición número 68 sin importar por cuál de los dos lados se observe (de abajo hacia arriba o de arriba hacia abajo), por tal motivo este dato divide en dos partes al conjunto de datos. Este dato se le conoce como **mediana**.

La **mediana** (\tilde{X}), es la medida de tendencia central que divide exactamente los datos en dos partes iguales después que estos han sido ordenados en forma ascendente o descendente.

La mitad de las observaciones es menor o igual que este valor y la otra mitad es mayor. en un conjunto de datos ordenados $x_1 \leq x_2 \leq x_3 \leq \dots \leq x_n$ la mediana se calcula teniendo en cuenta dos casos:

Caso 1: Si el número de observaciones “ n ” es impar:

$$\tilde{X} = x_{\left(\frac{n+1}{2}\right)}$$

Ejemplo:

Los siguientes datos representan el número de personas que asistieron al estreno de una película en su noche de estreno en las diferentes funciones y teatros de la ciudad de Bogotá.

43 42 46 43 44 32 47 48 49 45 41 40 42 50 55

Solución

Para determinar el valor de la mediana primero se organizan los datos de menor a mayor, así: $X_1 = 32$, $X_2 = 40$, $X_3 = 41$, $X_4 = 42$, $X_5 = 42$, $X_6 = 43$, $X_7 = 43$, $X_8 = 44$, $X_9 = 45$, $X_{10} = 46$, $X_{11} = 47$, $X_{12} = 48$, $X_{13} = 49$, $X_{14} = 50$, $X_{15} = 55$

Como el número de datos es $n = 15$ impar, entonces:

$$\tilde{X} = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{15+1}{2}\right)} = x_{\left(\frac{16}{2}\right)} = x_8 = 44$$

Luego, al 50 % de los teatros asistieron más de 44 personas y en el otro 50 % asistieron menos de 44 personas.

Caso 2: Si el número de observaciones “ n ” es par, la mediana es el punto medio entre las dos observaciones que se ubican en el centro.

$$\tilde{X} = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$$

Ejemplo:

Los siguientes puntajes corresponden a los resultados obtenidos por 20 estudiantes en las pruebas Icfes en el área de matemáticas.

32 32 33 34 36 38 40 45 46 48
49 49 50 51 52 56 58 60 61 64

1. Calcular la mitad de los puntajes obtenidos por los estudiantes.

Solución

Los datos se encuentran organizados de manera escendente, lo cual facilita el calculo de la mediana.

Como el número de datos es $n = 20$ par, entonces:

$$\tilde{X} = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} = \frac{x_{(\frac{20}{2})} + x_{(\frac{20}{2}+1)}}{2} = \frac{x_{10} + x_{11}}{2} = \frac{48 + 49}{2} = \frac{97}{2} = 48,5$$

Se puede concluir que la mitad de los estudiantes obtuvieron un resultado en la prueba menor o igual a 48,5 puntos y la otra mitad un resultado mayor a 48,5 puntos.

3.1.3. Moda

Los siguientes datos corresponden al número de hijos que tienen 15 familias escogidas al azar en la ciudad de Neiva: 2, 3, 4, 3, 3, 0, 3, 1, 2, 3, 1, 3, 4, 3, 3.

Facilmente se observa que hay un número (3) que se repite varias veces. Este dato recibe el nombre de **Moda**.

La **moda** (\hat{X}) de un conjunto de datos, corresponde al dato que más se repite, es decir, aquel dato que tiene mayor frecuencia.

La moda es la única medida de tendencia central que se puede utilizar en variable cualitativa. Veamos:

Ejemplo:

En la siguiente tabla se presenta el lugar de nacimiento de los pasajeros de un bus intermunicipal

Neiva	Campoalegre	Rivera	Campoalegre	Aipe	Neiva
Campoalegre	Palermo	Neiva	Palermo	Neiva	Aipe
Palermo	Neiva	Neiva	Palermo	Neiva	Neiva
Palermo	Yaguara	Neiva	Aipe	Palermo	Neiva

Solución

Como se puede observar el lugar de nacimiento que predomina es Neiva.

Para calcular el lugar promedio de nacimiento, no se puede por que se está trabajando con palabras y no con números.

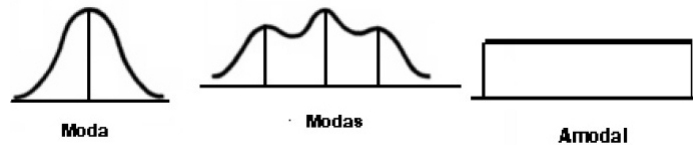
Una serie de datos puede tener uno, dos o más datos que se repiten a la vez, por tanto, una distribución puede ser *amodal*, *unimodal*, *bimodal* o *multimodal*.

Una distribución es **amodal** cuando no tiene moda, es decir, todas las observaciones ocurren con igual frecuencia.

Una distribución es **unimodal** cuando posee una sola moda.

Una distribución es **bimodal** cuando dos datos se repiten igual número de veces (dos modas).

Una distribución es **multimodal** cuando tres o más datos se repiten igual número de veces (tres o más modas).



Ejemplo:

A las primeras 12 personas que salieron de un banco en el centro de la ciudad, se les pregunto qué opinión tenían de la atención recibida, los resultados fueron los siguientes:

Buena Mala Excelente Excelente Mala Buena
Buena Excelente Buena Excelente Buena Excelente

Como se observa hay dos opiniones que predominan en las personas, la primera es buena con 6 respuestas y la segunda excelente también con 6 respuestas, debido a que las dos tienen igual cantidad de respuestas y son las mayores luego se dice que es bimodal.

3.2. Medidas de posición para datos no agrupados

Son números que dividen el conjunto de datos en partes iguales y se usan para clasificar una observación dentro de una población o muestra.

Estas medidas se utilizan para describir la posición que tiene un subconjunto de datos, con relación al resto de datos.

Dentro de estas medidas se encuentran: los *cuartiles*, *deciles* y *percentiles*.

3.2.1. Cuartiles

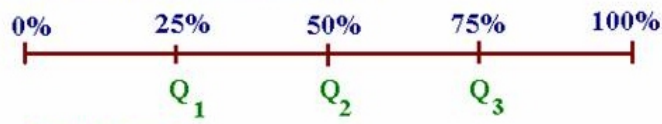
Los **cuartiles** (Q_1, Q_2, Q_3) son los valores de la variable, que dividen un conjunto ordenado de datos en cuatro partes iguales, por tanto, hay 3 valores que representan el 25 %, el 50 % y el 75 % de los datos.

Q_1 (valor que deja por debajo el **25 %** de los datos y por encima el **75 %** restante.)

Q_2 (coincide con la mediana, es decir, valor que deja por debajo el **50 %** de los datos y por encima el otro **50 %**.)

Q_3 (valor que deja por debajo el **75 %** de los datos y por encima el **25 %** restante.)

Para calcular los cuartiles se debe encontrar la mediana (\tilde{X}), es decir, Q_2 , y con los datos menores o iguales a Q_2 se calcula una nueva mediana que corresponde a Q_1 , y con los datos mayores o iguales a Q_2 se calcula otra nueva mediana que corresponde a Q_3 .



3.2.2. Deciles

Los **deciles** ($D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9$) son valores de los datos que divide el conjunto de datos en 10 partes iguales, cada uno representa el **10 %** de la distribución de los datos.

D_1 (valor que deja por debajo el **10 %** de los datos y por encima el **90 %** restante.)

D_2 (valor que deja por debajo el **20 %** de los datos y por encima el **80 %** restante.)

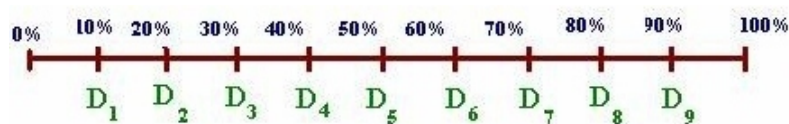
D_3 (valor que deja por debajo el **30 %** de los datos y por encima el **70 %** restante.)

D_4 (valor que deja por debajo el **40 %** de los datos y por encima el **60 %** restante.)

D_5 (coincide con la mediana, es decir, valor que deja por debajo el **50 %** de los datos y por encima el otro **50 %**.)

⋮ ⋮ ⋮

D_9 (valor que deja por debajo el **90 %** de los datos y por encima el **10 %** restante.)



3.2.3. Percentiles

Los **percentiles** ($P_1, P_2, P_3, \dots, P_{99}$) son valores de la variable que dividen el conjunto de datos ordenados en 100 partes iguales.

P_1 (valor que deja por debajo el **1 %** de los datos y por encima el **99 %** restante.)

P_2 (valor que deja por debajo el **2 %** de los datos y por encima el **98 %** restante.)

⋮ ⋮ ⋮

P_{25} (valor que deja por debajo el **25 %** de los datos y por encima el **75 %** restante.)

⋮ ⋮ ⋮

P_{50} (coincide con la mediana, es decir, valor que deja por debajo el **50 %** de los datos y por encima el **50 %** restante.)

⋮ ⋮ ⋮

P_{99} (valor que deja por debajo el **99 %** de los datos y por encima el **1 %** restante.)

Para un número de n observaciones una vez ordenados los datos, se puede identificar la posición de los cuartiles, deciles y porcentiles de acuerdo a las siguientes formulas:

$$n \text{ par} \implies \frac{k \times n}{A}$$

$$n \text{ impar} \implies \frac{k \times (n + 1)}{A}$$

Donde:

▸ **Cuartil:** k número del cuartil, es decir, 1, 2 o 3.

A número de divisiones del conjunto de datos y $A = 4$.

▸ **Decil:** k número del decil, es decir, 1, 2, 3, ..., 9.

A número de divisiones del conjunto de datos y $A = 10$.

▸ **Percentil:** k número del percentil cuartil, es decir, 1, 2, 3, ..., 99.

A número de divisiones del conjunto de datos y $A = 100$.

Existe un valor en el cual coinciden los *cuartiles*, los *deciles* y *percentiles*, y es cuando son iguales a la mediana, es decir, $Q_2 = D_5 = P_{50} = \widetilde{X}$. Todos dejan por debajo el 50 % de los datos y por encima el otro 50 %.

Ejemplo:

Una empresa que fabrica bombillos selecciono 10 de ellos al azar para realizar una prueba de duración en horas. Los datos que se obtuvieron fueron los siguientes:

5800	5890	5978	5998	6000	6213	6302	6529	6540	6541
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}

El fabricante quiere determinar cuanto tiempo de duración registran el 25 %, 50 %, 93 % de las bombillas que produce.

Solución

☞ Para determinar el 25 %, se halla la posición del cuartil número 1 (Q_1):

$$\frac{k \times n}{A} = \frac{1 \times 10}{4} = \frac{10}{4} = 2,5 \approx 3$$
, es decir, Q_1 es el dato de la posición 3, para este caso el cuartil $Q_1 = 5978$. Luego el 25 % de las bombillas registran un tiempo de duración menores o iguales a 5978 horas.

☞ Para determinar el 50 %, se halla la posición del decil número 5 (D_5).

$$\frac{k \times n}{A} = \frac{5 \times 10}{10} = \frac{50}{10} = 5$$
, es decir, D_5 es el dato de la posición 5, para este caso el decil $D_5 = 6000$. Luego el 50 % de las bombillas registran un tiempo de duración menores o iguales a 6000 horas y el 50 % restante mayor a 6000 horas.

☞ Para determinar el 93 %, se halla la posición del percentil número 93 (P_{93}).

$$\frac{k \times n}{A} = \frac{93 \times 10}{100} = \frac{930}{100} = 9,3 \approx 9$$
, es decir, P_{93} es el dato de la posición 9, para este caso el percentil $P_{93} = 6540$. Luego el 93 % de las bombillas registran un tiempo de duración menores o iguales a 6540 horas y el 7 % restante mayor a 6540 horas de duración.

3.3. Medidas de dispersión para datos no agrupados

La descripción de un conjunto no es completa citando únicamente las medidas de tendencia central. Esto puede observarse cuando se comparan dos muestras que tienen promedios muy similares.

Las *medidas de dispersión*, son valores numéricos que miden la dispersión, variabilidad o alejamiento entre los datos. Permiten conocer el grado de agrupamiento de los datos.

Si los datos están relativamente cerca unos de otros, con respecto a la escala en la cual se midieron, las medidas de dispersión toman valores pequeños. Si por el contrario los datos están relativamente lejanos unos de otros, las medidas de dispersión toman valores numéricos grandes. Las principales medidas de dispersión son: el *rango*, la *varianza*, la *desviación estándar* y el *coeficiente de variación*.

3.3.1. Rango

El **rango** (R) es la diferencia entre el dato más alto y el dato más bajo.

$$Rango = dato\ mayor - dato\ menor$$

Por medio del *rango* se puede conocer el campo de variación de los datos de la muestra, entre más pequeño sea el rango más cercanos estarán los datos.

Ejemplo:

A continuación se muestran las edades en años de las 15 primeras personas que fueron a consulta con el médico el día 30 de noviembre.

35 38 39 42 43 47 50 52 55 59 60 61 63 64 65

Calcular el recorrido de las edades de los pacientes.

Solución

Hallar el recorrido de las edades es lo mismo que calcular el rango en el cual está variando la característica descrita, es decir:

$$Rango = dato\ mayor - dato\ menor = 65 - 35 = 30$$

3.3.2. Varianza

La **varianza**, permite calcular el promedio de las diferencias al cuadrado entre el valor de cada dato y la media aritmética. Se simboliza S^2 y se calcula con la siguiente expresión:

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

La varianza presenta dos problemas, cuando se realiza su interpretación:

- Es un número muy grande con respecto a las observaciones, y
- Se expresa en términos de los datos originales elevados al cuadrado, los cuales no tienen una interpretación lógica, por ejemplo si la variable está dada en años, la varianza queda en años al cuadrado (no tiene sentido).

Estas dificultades se solucionan usando la **desviación estándar**.

3.3.3. Desviación estándar

La **desviación estándar**, permite medir la dispersión de los datos respecto al valor de la media o promedio; cuanto más grande sea su valor, más dispersos estarán los datos de la media. Se simboliza S y se halla como la raíz cuadrada positiva de la varianza, es decir:

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

3.3.4. Coeficiente de variación

El **coeficiente de variación** es una medida útil para comparar las dispersiones en varios conjuntos que tienen distintas escalas de medición o distintas unidades. Se simboliza **CV** y se calcula con la expresión:

$$CV = \frac{S}{\bar{X}}$$

Entre más pequeño sea el valor de la desviación estándar el coeficiente de variación mejor representa la media aritmética al conjunto de datos, en otras palabras, los datos se encuentran muy cerca del promedio.

Ejemplo:

La siguiente tabla de datos representa el puntaje obtenido por Luis en la prueba ICFES.

PRUEBA	BIOLOGÍA	MATEMÁTICAS	FILOSOFÍA	FÍSICA	HISTORÍA
PUNTAJE	59	90	58	60	52

QUÍMICA	LENGUAJE	GEOGRAFÍA	INGLES
60	68	62	68

Calcular

1. El promedio general del puntaje obtenido por Luis en la prueba Icfes.
2. La desviación típica y el coeficiente de variación.

Solución

1. El promedio del puntaje se calcula de la siguiente manera:
-

$$\bar{X} = \frac{59 + 90 + 58 + 60 + 52 + 60 + 68 + 62 + 68}{9} = \frac{577}{9} \approx 64,11 \text{ puntos}$$

2. La varianza (S^2)

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n} = \frac{\sum_{i=1}^9 (x_i - 64,11)^2}{9} = \frac{948,89}{9} = 105,43$$

Luego, $S = \sqrt{105,43} = 10,26$, entonces el promedio de alejamiento entre los puntajes y la \bar{X} es de 10,26.

Para calcular el Coeficiente de variación:

$$CV = \frac{S}{\bar{X}} = \frac{10,27}{64,11} = 0,160$$

El promedio de Luis fue de 64,11 puntos y los puntajes no se encuentran muy dispersos con respecto a la media.

Ejemplo:

A continuación se muestra el precio del dólar al momento del cierre de la jornada bursátil, durante el transcurso de 30 días.



De acuerdo a los datos del precio del dólar, se pide calcular:

1. El precio promedio del dolar.
2. El precio del dolar ubicado en la mitad.
3. El precio del dolar que más se repite.
4. Por debajo de que precio se encuentra el 75 % de los precios del dolar.

5. Por encima de que precio se encuentra el 60 % de los precios del dolar.
6. Por debajo de que precio se encuentra el 69 % de los precios del dolar.
7. Rango.
8. Varianza.
9. Desviación estándar.
10. Coeficiente de variación.

fecha	cierre	fecha	cierre	fecha	cierre
19 - junio - 2010	1906,61	29 - junio - 2010	1916,46	9 - julio - 2010	1877,66
20 - junio - 2010	1906,61	30 - junio - 2010	1913,15	10 - julio - 2010	1877,66
21 - junio - 2010	1889,51	1 - julio - 2010	1897,33	11 - julio - 2010	1877,66
22 - junio - 2010	1886,05	2 - julio - 2010	1884,31	12 - julio - 2010	1877,20
23 - junio - 2010	1895,75	3 - julio - 2010	1884,31	13 - julio - 2010	1871,19
24 - junio - 2010	1896,87	4 - julio - 2010	1884,31	14 - julio - 2010	1873,87
25 - junio - 2010	1900,36	5 - julio - 2010	1884,31	15 - julio - 2010	1871,96
26 - junio - 2010	1900,36	6 - julio - 2010	1882,47	16 - julio - 2010	1878,77
27 - junio - 2010	1900,36	7 - julio - 2010	1901,38	17 - julio - 2010	1878,77
28 - junio - 2010	1901,65	8 - julio - 2010	1889,61	18 - julio - 2010	1878,77

Solución

Para facilitar el cálculo de las medidas estadísticas, se organizan los datos de manera ascendente, con lo cual se obtiene:

$X_1 = 1871,19$	$X_7 = 1877,66$	$X_{13} = 1884,31$	$X_{19} = 1895,75$	$X_{25} = 1901,38$
$X_2 = 1871,96$	$X_8 = 1878,77$	$X_{14} = 1884,31$	$X_{20} = 1896,87$	$X_{26} = 1901,65$
$X_3 = 1873,87$	$X_9 = 1878,77$	$X_{15} = 1884,31$	$X_{21} = 1897,33$	$X_{27} = 1906,61$
$X_4 = 1877,20$	$X_{10} = 1878,77$	$X_{16} = 1886,05$	$X_{22} = 1900,36$	$X_{28} = 1906,61$
$X_5 = 1877,66$	$X_{11} = 1882,47$	$X_{17} = 1889,51$	$X_{23} = 1900,36$	$X_{29} = 1913,15$
$X_6 = 1877,66$	$X_{12} = 1884,31$	$X_{18} = 1889,61$	$X_{24} = 1900,36$	$X_{30} = 1916,46$

1. Media aritmética (\bar{X})

Para calcular la media aritmética, se suman todos los precios del dólar y el resultado se divide entre el total de precios ($n = 30$), es decir,

$$\bar{X} = \frac{\sum_{i=1}^{30} x_i}{30} = \frac{56685,28}{30} = 1889,509 \text{ pesos}$$

Luego, el precio promedio del dolar durante estos 30 días es 1889,509 pesos.

2. Mediana (\tilde{X})

Como el número de observaciones ($n = 30$) es par, entonces se utiliza la formula para *n par*:

$$\begin{aligned} \tilde{X} &= \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2})+1}}{2} = \frac{x_{(\frac{30}{2})} + x_{(\frac{30}{2})+1}}{2} = \frac{x_{15} + x_{16}}{2} \\ &= \frac{1884,31 + 1886,05}{2} = \frac{3770,36}{2} = 1885,18 \text{ pesos} \end{aligned}$$

Se puede concluir que la mitad de los precios del dolar son menores o iguales a 1885,18 pesos y la otra mitad un precio mayor a 1885,18 pesos.

3. Moda.

En este conjunto de observaciones, el dato que más se repite es 1885,31 pesos debido a que se repite 4 veces. ($X_{12}, X_{13}, X_{14}, X_{15}$)

4. Tercer cuartil (Q_3)

$$\frac{k \times n}{A} = \frac{3 \times 30}{4} = \frac{90}{4} = 22,5 \approx 23$$

Es decir, Q_3 es el dato de la posición 23, en este caso el cuartil $Q_3 = 1900,36$. Luego el 75% de los precios del dolar es menor o igual a \$ 1900,36 y el 25% restante mayor a \$ 1900,36.

5. Cuarto decil (D_4)

$$\frac{k \times n}{A} = \frac{4 \times 30}{10} = \frac{120}{10} = 12$$

Como 12 es un dato exacto, para hallar el decil 4 se promedian el dato de la posición 12 y el dato de la posición 13.

$$D_4 = \frac{X_{12} + X_{13}}{2} = \frac{1884,31 + 1884,31}{2} = 1884,31$$

Luego el 40% de los precios del dolar fueron menores o iguales a \$ 1884,31 y el 60%, mayores \$ 1884,31.

6. Percentil 31 (P_{31})

$$\frac{k \times n}{A} = \frac{31 \times 30}{100} = \frac{930}{100} = 9,3 \approx 9$$

Es decir, P_{31} es el dato de la posición 9, en este caso el percentil $P_{31} = 1878,77$. Luego el 31% de los precios del dolar es menor o igual a \$ 1878,77 y el 69% restante mayor a \$ 1878,77.

7. Rango

$$\text{Rango} = X_{30} - X_1 = 1916,46 - 1871,19 = 45,27$$

Luego, la variación de los precios del dolar se encuentran en un rango de 45,27 pesos.

8. Varianza (S^2)

$$\begin{aligned} S^2 &= \frac{\sum_{i=1}^{30} (x_i - \bar{X})^2}{30} = \frac{\sum_{i=1}^{30} (x_i - 1889,509)^2}{30} \\ &= \frac{4644,46}{30} = 154,82 \text{ pesos} \end{aligned}$$

9. Desviación estándar (S)

$$S = \sqrt{S^2} = \sqrt{154,82} \approx 12,44 \text{ pesos}$$

En conclusión, se puede decir que los precios del dolar no están tan dispersos con respecto al precio promedio.

10. Coeficiente de variación (CV)

$$CV = \frac{S}{\bar{X}} = \frac{12,44}{1889,509} = 0,00658$$

3.4. Medidas de tendencia central para datos agrupados

Al igual que los datos sin agrupar, se pueden determinar medidas estadísticas cuando los datos se encuentran agrupados con o sin intervalos con el fin de determinar el comportamiento de las observaciones.

3.4.1. Media aritmética

La media (\bar{X}) en datos agrupados se calcula utilizando la siguiente fórmula:

$$\bar{X} = \frac{\sum_{i=1}^n (x_i \times f_i)}{n}$$

Donde:

⇒ n : total de datos.

⇒ x_i : dato, o la marca de clase.

⇒ f_i : frecuencia del dato.

⇒ $\sum_{i=1}^n$: sumatoria.

Ejemplo:

En una liga de fútbol se realizó un estudio sobre las anotaciones hechas en una temporada. Los resultados se muestran en la siguiente tabla y se pide determinar el promedio de goles por partidos.

Número de anotaciones	f_i
0	2
1	14
2	9
3	6
4	7
5	2
	40

Solución

Para hallar el promedio de goles por partido durante la temporada, se utiliza la siguiente formula:

$$\begin{aligned}\bar{X} &= \frac{\sum_{i=1}^n (x_i \times f_i)}{n} = \frac{(0 \times 2) + (1 \times 14) + (2 \times 9) + (3 \times 6) + (4 \times 7) + (5 \times 2)}{40} \\ &= \frac{0 + 14 + 18 + 18 + 28 + 10}{40} = \frac{88}{40} = 2,2 \text{ es decir, en promedio durante la liga se marcaron } \\ &2 \text{ goles por partido.}\end{aligned}$$

Ejemplo:

Una clínica tiene una tabla donde registra las edades de sus pacientes. En la tabla adjunta se presenta la información de los pacientes atendidos durante una semana.

Determinar la edad promedio de los pacientes que acudieron a la clínica durante esa semana

Edad (Años)	m_i	f_i
[10 - 20)	15	8
[20 - 30)	25	20
[30 - 40)	35	14
[40 - 50)	45	8
[50 - 60)	55	2
[60 - 70)	65	2
[60 - 70)	75	1
	Total	55

Solución

Para saber cuál es la edad promedio de los pacientes atendidos durante una semana:

$$\begin{aligned}\bar{X} &= \frac{\sum_{i=1}^n (m_i \times f_i)}{n} \\ &= \frac{(15 \times 8) + (25 \times 20) + (35 \times 14) + (45 \times 8) + (55 \times 2) + (65 \times 2) + (75 \times 2)}{55} \\ &= \frac{120 + 500 + 490 + 360 + 110 + 130 + 75}{55} = \frac{1785}{55} = 32,4\end{aligned}$$

es decir, la edad promedio de los pacientes que acudieron durante una semana es de 32 años.

3.4.2. Mediana

La mediana (\tilde{X}) en datos agrupados para variable cuantitativa discreta se calcula utilizando las mismas fórmulas que se emplean para calcularla en datos no agrupados, y se tienen en cuenta los mismos dos casos.

La mediana (\tilde{X}) en datos agrupados con intervalos, como primer paso se debe identificar el intervalo en donde está la mediana y se aplica la siguiente fórmula:

$$\tilde{X} = L_i + C_k \left(\frac{\frac{n}{2} - F_{k-1}}{f_k} \right)$$

Donde:

- ⇒ n : Total de datos
- ⇒ L_i : Límite inferior del intervalo.
- ⇒ C_k : Amplitud del intervalo.
- ⇒ F_{k-1} : Frecuencia absoluta acumulada del intervalo anterior.
- ⇒ f_k : Frecuencia absoluta del intervalo.

Para identificar el intervalo en donde se encuentra la mediana se cuentan el número de observaciones del problema, si es par se ubican los dos datos del centro, pero si es impar se busca el dato central. Después se ubica el intervalo al cual pertenece el o los datos.

3.4.3. Moda

La moda (\widehat{X}) en datos agrupados, se identifica con la mayor frecuencia absoluta f , cuando la variable es cuantitativa discreta.

La moda (\widehat{X}) en datos agrupados por intervalos, primero procede a identificar el intervalo con mayor frecuencia absoluta f y luego se aplica la siguiente fórmula:

$$\widehat{X} = L_i + C_k \left(\frac{f_k - f_{k-1}}{2f_k - f_{k-1} - f_{k+1}} \right)$$

Donde:

- ⇒ L_i : Límite inferior del intervalo.
- ⇒ C_k : Amplitud del intervalo.
- ⇒ f_{k-1} : Frecuencia absoluta del intervalo anterior.
- ⇒ f_k : Frecuencia absoluta del intervalo.
- ⇒ f_{k+1} : Frecuencia absoluta del intervalo siguiente.

3.5. Medidas de posición para datos agrupados

3.5.1. Cuartil, Decil y Percentil

Para datos agrupados, se puede determinar la posición del percentil, el decil o el cuartil que se necesita simplemente aplicando la siguiente fórmula:

$$\frac{t \times n}{100}$$

Donde:

- ⇒ n : número total de datos.
-

⇒ t : porcentaje que se quiere encontrar, así:

⇒ **Cuartiles** $t = 25, 50$ o 75

⇒ **Deciles** $t = 10, 20, 30, \dots, 100$

⇒ **Percentiles** $t = 1, 2, 3, 4, \dots, 100$

Para datos agrupados por intervalos, se identifica en cual intervalo se encuentra el dato utilizado la formula anterior y se aplica:

$$L_i + C_k \left(\frac{0,01t \times n - F_{k-1}}{f_k} \right)$$

Donde:

⇒ t : porcentaje que se quiere encontrar, así.

⇒ **Cuartiles** $t = 25, 50$ o 75

⇒ **Deciles** $t = 10, 20, 30, \dots, 100$

⇒ **Percentiles** $t = 1, 2, 3, 4, \dots, 100$

⇒ n : Total de datos.

⇒ f_k : Frecuencia absoluta del intervalo.

⇒ F_{k-1} : Frecuencia absoluta acumulada del intervalo anterior.

Ejemplo:

El consumo semanal de pescado en una población ribereña del río Magdalena se muestra en la siguiente tabla:

1. Determinar que cantidad de pescado consume:

a. El 60% de la población.

b. El 75% de la población.

Consumo (Kg./Sem)	f_i	F_i	m_i
[0 - 1,5)	15	15	0,75
[1,5 - 3,0)	26	41	2,25
[3,0 - 4,5)	20	61	3,75
[4,5 - 6,0)	13	74	5,25
[6,0 - 7,5)	6	80	6,75
Total	80		

c. El 85 % de la población.

Solución

a. Para determinar la cantidad de pescado que consume el 60 % de la población, se halla el decil número 6 (D_6).

Para ello se identifica el intervalo en el cual se encuentra el dato, esto se logra utilizando $\frac{t \times n}{100}$, donde $t = 60$, $n = 80$.

$$\frac{t \times n}{100} = \frac{60 \times 80}{100} = \frac{4800}{100} = 48$$

El dato 48 pertenece al intervalo [3,0 - 4,5), luego se aplica:

$$D_6 = L_i + C_k \left(\frac{(0,01t \times n) - F_{k-1}}{f_k} \right)$$

Donde:

$$L_i = 3,0 \quad C_k = 1,5 \quad t = 60 \quad n = 80 \quad F_{k-1} = 41 \quad f_k = 20$$

Se tiene:

$$D_6 = 3,0 + 1,5 \times \left(\frac{(0,01 \times 60 \times 80) - 41}{20} \right) = 3,0 + 1,5 \times \left(\frac{7}{20} \right)$$

$$= 3,0 + 0,53 = 3,53 \text{ Kilogramos}$$

Entonces, el 60 % de la población encuestada consume 3,53 Kg. o menos de pescado.

- b. Para determinar la cantidad de pescado que consume el 75 % de la población, se halla el cuartil número 3 (Q_3).

Para ello identificamos el intervalo en el cual se encuentra el dato, esto se logra utilizando $\frac{t \times n}{100}$, donde $t = 75$, $n = 80$.

$$\frac{t \times n}{100} = \frac{75 \times 80}{100} = \frac{6000}{100} = 60$$

El dato 60 pertenece al intervalo $[3,0 - 4,5)$, luego:

$$Q_3 = L_i + C_k \left(\frac{0,01t \times n - F_{k-1}}{f_k} \right)$$

donde:

$$L_i = 3,0 \quad C_k = 1,5 \quad t = 75 \quad n = 80 \quad F_{k-1} = 41 \quad f_k = 20$$

Se tiene:

$$Q_3 = 3,0 + 1,5 \times \left(\frac{(0,01 \times 75 \times 80) - 41}{20} \right) = 3,0 + 1,5 \times \left(\frac{19}{20} \right)$$

$$= 3,0 + 1,425 = 4,425 \text{ Kilogramos}$$

Luego, el 75 % de la población encuestada consume 4,425 Kg. o menos de pescado.

- c. Para determinar la cantidad de pescado que consume el 85 % de la población, se halla el percentil número 85 (P_{85}).

Para ello se identifica el intervalo en el cual se encuentra el dato, esto se logra utilizando $\frac{t \times n}{100}$, donde $t = 85$, $n = 80$.

$$\frac{t \times n}{100} = \frac{85 \times 80}{100} = \frac{6800}{100} = 68$$

El dato 68 pertenece al intervalo $[4, 5 - 6, 0)$, luego:

$$P_{85} = L_i + C_k \left(\frac{0,01t \times n - F_{k-1}}{f_k} \right)$$

Donde:

$$L_i = 4,5 \quad C_k = 1,5 \quad t = 85 \quad n = 80 \quad F_{k-1} = 61 \quad f_k = 13$$

Se tiene:

$$\begin{aligned} P_{85} &= 4,5 + 1,5 \times \left(\frac{(0,01 \times 85 \times 80) - 61}{13} \right) = 4,5 + 1,5 \times \left(\frac{7}{13} \right) \\ &= 4,5 + 0,807 = 5,30 \text{ Kilogramos} \end{aligned}$$

Entonces, el 85 % de la población encuestada consume 5,30 Kg. o menos de pescado.

3.6. Medidas de dispersión para datos agrupados

3.6.1. Varianza

La varianza (S^2) se calcula en datos agrupados, utilizando la siguiente fórmula:

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2 f_i}{(n-1)}$$

NOTA: Cuando los datos están agrupados por intervalos x_i es la marca de clase de cada intervalo.

3.6.2. Desviación estándar

La desviación estándar (S), se halla para datos agrupados, obteniendo la raíz cuadrada positiva de la varianza (S^2) de datos agrupados:

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2 f_i}{(n - 1)}}$$

3.6.3. Coeficiente de variación

El coeficiente de variación (CV) se calcula para datos agrupados y datos no agrupados con intervalos, de la misma forma como se calcula para datos no agrupados, es decir, aplicando la siguiente fórmula:

$$CV = \frac{S}{\bar{X}}$$

Ejemplo:

Desde el pasado 27 de agosto de 2007, los colombianos tuvieron la opción de adquirir las acciones de Ecopetrol, las cuales los habilitan como nuevos accionistas de la primera empresa del país ranqueada entre las cuarenta principales petroleras del mundo.



Cada acción de Ecopetrol al iniciar su venta tenía un valor de \$ 1.400, y lo mínimo que se podía adquirir eran 1.000 acciones. Ecopetrol es hoy la cuarta petrolera de América Latina.

Desde el momento en que se inicio la venta de las acciones, estas han tenido una valorización muy considerable en la bolsa accionaria colombiana. El precio de la acción de Ecopetrol al momento del cierre durante el mes de junio del año 2010 se presenta en la siguiente tabla:

fecha	cierre	fecha	cierre	fecha	cierre
1 - junio - 2010	2730	11 - junio - 2010	2710	23 - junio - 2010	2760
2 - junio - 2010	2730	15 - junio - 2010	2720	24 - junio - 2010	2770
3 - junio - 2010	2720	16 - junio - 2010	2750	25 - junio - 2010	2775
4 - junio - 2010	2690	17 - junio - 2010	2750	28 - junio - 2010	2750
8 - junio - 2010	2685	18 - junio - 2010	2765	29 - junio - 2010	2740
9 - junio - 2010	2700	21 - junio - 2010	2750	30 - junio - 2010	2755
10 - junio - 2010	2710	22 - junio - 2010	2740		

Organizados los precios de la acción en una tabla de valores con intervalos:

Precio de acción de Ecopetrol	<i>m</i>	<i>f</i>	<i>F</i>
[2685 - 2703)	2694	3	3
[2703 - 2721)	2712	4	7
[2721 - 2739)	2730	2	9
[2739 - 2757)	2748	7	16
[2757 - 2775)	2766	4	20

De acuerdo a los precios de la acción de Ecopetrol registrados durante el mes de Junio, calcular:

1. El precio promedio de la acción.
2. El precio localizado en la mitad.
3. El precio que más se repitió.

4. El 25 % de los precios de la acción de Ecopetrol, por debajo de que precio se encuentran?
5. El 42 % de los precios de la acción de Ecopetrol, cuál precio máximo registró?
6. Varianza.
7. Desviación estándar.
8. Coeficiente de variación.

Solución

1. Media aritmética (\bar{X})

$$\begin{aligned}\bar{X} &= \frac{\sum_{i=1}^n (m_i f_i)}{n} \\ &= \frac{(2694 \times 3) + (2712 \times 4) + (2730 \times 2) + (2748 \times 7) + (2766 \times 4)}{30} \\ &= \frac{54690}{30} = 2734,5 \text{ pesos}\end{aligned}$$

Luego, durante estos días de cotización el precio promedio de la acción de Ecopetrol fue de 2734,5 pesos.

2. Mediana (\tilde{X})

Como el número de observaciones es par ($n = 20$), se ubican los dos datos del centro, es decir, (X_{10}, X_{11}) los cuales pertenecen al intervalo $[2739 - 2757)$. Aplicando la siguiente formula:

$$\tilde{X} = L_i + C_k \left(\frac{0,5 n - F_{k-1}}{f_k} \right)$$

Donde:

$$L_i = 2739 \quad C_k = 18 \quad n = 20 \quad F_{k-1} = 9 \quad f_k = 7$$

Tenemos:

$$\begin{aligned}\tilde{X} &= 2739 + 18 \times \left(\frac{(0,5 \times 20) - 9}{7} \right) = 2739 + 18 \times \left(\frac{1}{7} \right) \\ &= 2739 + 2,57 = 2741,57 \text{ pesos}\end{aligned}$$

Es decir, el precio de la acción que se encuentra en toda la mitad es 2741,57 pesos.

3. Moda (\hat{X})

Se identifica el intervalo con mayor frecuencia absoluta (f), que es [2739 – 2757]. Con la ayuda de la siguiente formula se halla el valor de la moda:

$$\hat{X} = L_i + C_k \left(\frac{f_k - f_{k-1}}{2 f_k - f_{k-1} - f_{k+1}} \right)$$

Con:

$$L_i = 2739 \quad C_k = 18 \quad f_{k-1} = 2 \quad f_k = 7 \quad f_{k+1} = 4$$

Se tiene:

$$\begin{aligned}\hat{X} &= 2739 + 18 \times \left(\frac{7 - 2}{[(2 \times 7) - 2] - 4} \right) = 2739 + 18 \times \left(\frac{5}{8} \right) \\ &= 2739 + 11,25 = 2750,25 \text{ pesos}\end{aligned}$$

Entonces, se concluye que el valor que más se repitió fue 2750,25 pesos.

4. Cuartil 1 (Q_1)

Como se está hallando el 25 % entonces se calcula el Q_1 para ello se identifica el intervalo en el cual se encuentra el dato, esto se logra utilizando $\frac{t \times n}{100}$, donde $t = 25$, $n = 20$.

$$\frac{t \times n}{100} = \frac{25 \times 20}{100} = \frac{500}{100} = 5$$

El dato 5 pertenece al intervalo $[2703 - 2721)$, luego:

$$Q_1 = L_i + C_k \left(\frac{0,01t \times n - F_{k-1}}{f_k} \right)$$

Donde:

$$L_i = 2703 \quad C_k = 18 \quad t = 25 \quad n = 20 \quad F_{k-1} = 3 \quad f_k = 3$$

Se tiene:

$$\begin{aligned} Q_1 &= 2703 + 18 \times \left(\frac{(0,01 \times 25 \times 20) - 3}{4} \right) = 2703 + 18 \times \left(\frac{2}{4} \right) \\ &= 2703 + 9 = 2712 \text{ pesos} \end{aligned}$$

Entonces, el 25 % de los precios de las acciones de Ecopetrol registran un precio menor o igual a 2712 pesos.

5. Percentil 42 (P_{42})

Se identifica el intervalo en el cual está el (P_{42}) utilizando la siguiente formula $\frac{t \times n}{100}$, donde $t = 42$, $n = 20$.

$$\frac{t \times n}{100} = \frac{42 \times 20}{100} = \frac{840}{100} = 8,4$$

El dato 8,4 está contenido en el intervalo $[2721 - 2739)$, con la ayuda de:

$$P_{42} = L_i + C_k \left(\frac{0,01t \times n - F_{k-1}}{f_k} \right)$$

Con:

$$L_i = 2721 \quad C_k = 18 \quad t = 42 \quad n = 20 \quad F_{k-1} = 7 \quad f_k = 2$$

Se tiene:

$$\begin{aligned} P_{42} &= 2721 + 18 \times \left(\frac{(0,01 \times 42 \times 20) - 7}{2} \right) = 2703 + 18 \times (0,7) \\ &= 2721 + 12,6 = 2733,6 \text{ pesos} \end{aligned}$$

Por consiguiente, el 42 % de los precios de la acción presenta un precio menor o igual a 2733,6 pesos y el 58 % restantes presentan un valor superior a 2733,6 pesos.

6. Varianza (S^2)

$$\begin{aligned} S^2 &= \frac{\sum_{i=1}^n (m_i - \bar{X})^2 f_i}{(n - 1)} \\ &= \frac{4920,75 + 2025 + 40,5 + 1215,75 + 3969}{19} \\ &= \frac{12231}{19} = 643,74 \end{aligned}$$

7. Desviación estándar (S)

$$S = \sqrt{S^2} = \sqrt{643,74} \approx 25,37$$

Luego, los precios de la acción de Ecopetrol se agrupan alrededor del precio promedio, debido a que la desviación estandar es pequeña, 25,37 dolares.

8. Coeficiente de variación (CV)

$$CV = \frac{S}{\bar{X}} = \frac{25,37}{2734,5} = 0,0092$$

3.7. Talleres de aplicaciones

1. La siguiente tabla muestra los datos de la población de un país en los últimos cinco años del siglo pasado:

Años	1995	1996	1997	1998	1999
Población	38541631	39295798	40064093	40826816	41589018

Calcular:

- Las medidas de centralización.
 - Las medidas de dispersión.
 - Las medidas de posición.
2. Determina qué valores se pueden agregar al siguiente grupo de datos para que la mediana siga siendo la misma.

34, 23, 35, 30, 32, 34, 32, 36, 35, 32, 29, 36

3. Agregar los datos necesarios al siguiente conjunto para que la media sea 11.

12, 13, 14, 11, 10, 11, 11, 14, 11, 15, 10, 12, 8

4. El promedio de notas de cinco estudiantes es 3,45. Si las notas de cuatro de ellos son: 3, 2; 4, 3; 3, 2; 2, 5; , ¿cuál es la quinta nota?
5. El promedio de seis datos es 85 y cinco de los datos son 20, 75, 120, 2, 4. ¿Cuál es el dato que falta?
6. En un conjunto de datos el coeficiente de variación es 0,8 y la desviación estándar es 1,5. ¿Cuál es el valor medio de los datos?
-

7. La siguiente tabla muestra las edades en años, de 45 estudiantes del grupo noveno del colegio “Neiva te quiero”

Años	No. de estudiantes
10	15
11	5
12	5
13	14
14	6

- a. Encontrar la edad promedio de un niño del grupo noveno.
- b. Hallar la mediana de las edades.
- c. Calcular los cuartiles.
- d. Hallar los 4 primeros deciles.
- e. Escribir varias conclusiones con respecto al grupo de noveno.
8. Los siguientes datos representan las edades de los 30 trabajadores de un hotel de la ciudad de Neiva.

22 20 20 19 21 20 18 27 23 19 21 19 30 20 21
55 29 27 22 25 30 25 21 28 23 22 24 22 21 25

- a. Encontrar la media y la mediana de las edades de los trabajadores.
- b. Calcular la media y la mediana, para las edades, sin tener en cuenta el trabajador cuya edad es 55 años. ¿Cambian las medidas?, ¿Cuál sufre un cambio mayor?, ¿A qué cree que se debe ese cambio?
-

- c. Hallar el primer cuartil, el quinto decil y el percentil sexto de las edades de los trabajadores.
9. En el proceso para fabricar baterías, se seleccionaron 10 de ellas para realizar una prueba de duración en horas. Los resultados fueron los siguientes:
- 52,5 58,7 55,3 56 54,8 57,9 55,8 52,5 54,8 54,7
- a. Hallar las medidas de tendencia central, los cuartiles 1 y 3, el decil 8 y el percentil 68.
10. Si la desviación estándar de un grupo de observaciones es 4,8, ¿cuál es la varianza?
11. La profesora de física del grado once analiza los resultados del último examen presentado por sus 22 alumnos.

Las notas son: 2, 7, 9, 6, 8, 4, 6, 1, 2, 7, 2, 9, 5, 1, 6, 6, 2, 9, 10, 4, 7, 5.

- a. Determina la nota promedio del curso en el área de física.
- b. Calcula la nota que se encuentra en toda la mitad.
- c. Cúal de las dos medidas anteriores representa mejor los datos. ¿Por qué?
- d. Escriba dos conclusiones sobre las notas de los estudiantes.
-

EXPERIMENTOS ALEATORIOS Y PROBABILIDAD

En la vida real existen experimentos tales como:

Exp. 1 ✓ Si se calienta el agua, ¿esta siempre ebullición a 100°C ?

Exp. 2 ✓ Si se deja caer un objeto desde una altura, éste siempre tardará el mismo tiempo en llegar al suelo?

Exp. 3 ✓ Si se lanza hacia arriba un objeto, se sabe que subirá hasta cierta altura, pero después bajará.

Exp. 4 ✓ $\text{H} + \text{H} + \text{O}$ produce agua.

Exp. 5 ✓ Si hoy es Martes, mañana será Miércoles.

Exp. 6 ✓ Si un número natural es par, el siguiente es impar.

Son experimentos que al ser repetidos en varias oportunidades bajo las mismas condiciones, sus resultados serán siempre los mismos.

Pero también existen otro tipo de experimentos:

Supongamos que se dispone de un dado con todas las caras pintadas de blanco y con un número, que irá de 1 a 6 sin repetir ninguno, en cada una de las seis caras. Definamos los dos experimentos siguientes:

Experimento 1: Tirar el dado y anotar el número de la cara resultante.



Experimento 2: Tirar nuevamente el dado y anotar el número de la cara resultante.

¿Qué diferencia fundamental observa entre los dos experimentos?

En el experimento 1, el resultado será una cara con un número cualquiera del 1 al 6 y lo mismo sucederá en el experimento 2, pero al repetir en varias oportunidades los experimentos no hay certeza de que los dados caigan en el mismo número, aunque dichas repeticiones sean realizadas bajo las mismas condiciones.

Otros ejemplos:

Experimento 3: Lanzar una moneda.

Experimento 4: Elegir al azar un estudiante de noveno grado.

Experimento 5: Extracción de una carta en una baraja.

Experimento 6: El próximo sorteo de lotería de Navidad, etc...

Luego,

Un experimento o fenómeno es aleatorio cuando al repetir el experimento en igualdad de condiciones los resultados varían, a pesar de mantener constantes las condiciones en las que este se realiza.

Espacio muestral:

En todo experimento aleatorio existe un conjunto de resultados posibles. Este conjunto recibe el nombre de espacio muestral del experimento. En otras palabras,

Es el conjunto formado por todos los posibles resultados de un experimento aleatorio. En adelante se designará por la letra **S**.

Ejemplos:

- ♥ Al lanzar un dado y se observa el resultado de la cara superior. El espacio muestral.
 $S = \{1, 2, 3, 4, 5, 6\}$
- ♥ Al lanzar al aire una moneda: $S = \{c, s\}$
- ♥ Al escoger un día de la semana $S = \{\text{Lunes, Martes, Miercoles, Jueves, Viernes, Sábado, Domingo}\}$.
- ♥ Al visitar un centro comercial de Neiva $S = \{\text{San Pedro Plaza, Metropolitano}\}$.
- ♥ Al lanzar 2 monedas simultáneamente: $S = \{(c,c) (c,s) (s,c) (s,s)\}$.
- ♥ Al tener cuatro personas, Carlos, Julio, Diana y Betty, únicamente para escoger al azar dos sin importar el orden: $S = \{(C,J) (C,D) (C,B) (J,D) (J,B) (D,B)\}$.
- ♥ Al lanzar 2 dados: $S = \{(1,2) (1,3) (1,4) (1,5) \dots (6,2) (6,3) (6,4) (6,5) (6,6)\}$

Evento o Suceso: Es un subconjunto del espacio muestral.

Ejemplos:

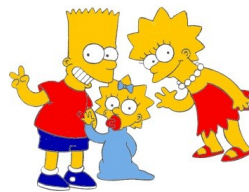
- ✿ Al caer cara cuando se lanza una moneda: $E = \{c\}$.
- ✿ Al caer 6 en un dado $\{6\}$.
- ✿ Cuando al lanzar dos dados la suma de sus puntos es un número par: $E = \{(1,1) (1,3) (1,5) (2,2) (2,4) (2,6) (3,1) (3,3) (3,5) (4,2) (4,4) (4,6) (5,1) (5,3) (5,5) (6,2) (6,4) (6,6)\}$.
- ✿ Cuando al lanzar 2 monedas, los resultados obtenidos son iguales: $E = \{(c,c) (s,s)\}$.
- ✿ Cuando al tener que escoger dos personas de 4 al azar (Carlos, Julio, Diana y Betty) y únicamente se quieren que sean del mismo sexo: $E = \{(C,J) (D,B)\}$.

- ✿ Un profesor debe seleccionar a tres estudiantes entre Julián, María, Vicente, Nidia y Paola, para un concurso de ortografía. Se pueden tener los siguientes eventos: Que sean seleccionadas solo mujeres, que sean seleccionados dos mujeres y un hombre, que en el grupo seleccionado esté un hombre, que Vicente no sea seleccionado.

Ejemplo:

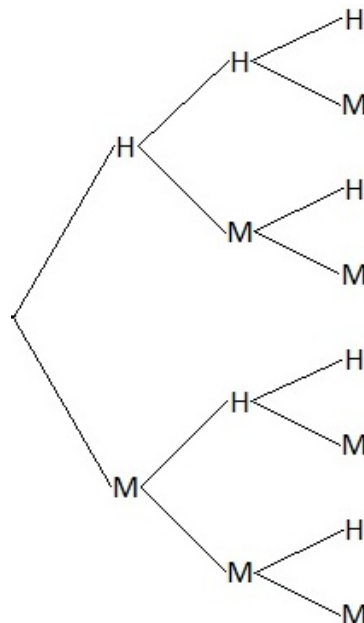
Una familia quiere tener tres hijos:

1. ¿Cuál es el experimento?
2. ¿Cuál es el espacio muestral?
3. Enumera 3 posibles eventos.



Solución

1. El experimento aleatorio consiste en que una familia quiere tener 3 hijos sin importar el orden en que puedan nacer.
2. El espacio muestral:
 $S = \{HHH, HHM, HMH, HMM, MHH, MHM, MMH, MMM\}$



3. Un evento puede ser que los tres hijos fueran del mismo sexo:

$$E_1 = \{(HHH) (MMM)\}$$

Otro evento, es cuando se quieren tener dos hijos del mismo sexo, en cualquier orden.

$$E_2 = \{(H H M) (H M H) (M H H) (M H M) (M M H) (H M M)\}$$

Y un tercer evento, cuando se quiere que el hijo mayor sea hombre

$$E_3 = \{(H H H) (H H M) (H M H) (H M M)\}$$

4.1. Reseña histórica de la Probabilidad

La creación de la Probabilidad se atribuye a los matemáticos franceses del siglo XVII Blaise Pascal y Pierre de Fermat, aunque algunos matemáticos anteriores, como Gerolamo Cardano en el siglo XVI, habían aportado importantes contribuciones a su desarrollo.

La Probabilidad matemática comenzó como un intento de responder a varias preguntas que surgían en los juegos de azar, por ejemplo, saber cuántos dados hay que lanzar para que la posibilidad de que salga algún seis supere el 50 %.



El concepto de *Probabilidad* puede ser interpretado como algo indefinible, pero utilizado para expresar, de algún modo, un “grado de creencia” que se tiene de la ocurrencia de un suceso; hace referencia a algo que puede suceder con base en la experiencia que se tenga. Cuando se escucha una expresión como “la probabilidad de que hoy llueva es del 70 %”, se está asignando una medida a la probabilidad.

Los hinchas de los diferentes equipos de fútbol discuten frecuentemente sobre la posibilidad de clasificar o de ganar un campeonato; el mismo alumno cuando se refiere a la posibilidad que tiene de ganar o de perder una asignatura; todos estos son pronósticos que se hacen con la esperanza de que sucedan.

Las probabilidades conjuntamente con la estadística tienen infinidad de aplicaciones a problemas de economía y ciencias sociales, de la misma manera a las ciencias físicas, industria, comercio y gobierno, con la observación de que en cada uno de ellos tendrá sus requisitos particulares.

Probabilidad

Definir adecuadamente la palabra *Probabilidad* es casi imposible, sin embargo hay alguna forma muy cercana al significado real de este término. En primer lugar se dirá que la probabilidad es un número comprendido entre 0 y 1.

$$0 \leq P \leq 1$$

Cuanto más cercano esté al valor a 1, tanto mayor será la posibilidad de que el evento suceda.

Si el valor de la probabilidad está muy cercano a 0, existe la posibilidad de que el evento no ocurra.

A medida que aumenta la cantidad de veces de un experimento aleatorio, la razón entre los éxitos y el total de repeticiones es cada vez más cercana a la probabilidad de dicho experimento.

Ejemplo:

Si se lanza una moneda 20 veces al aire, es posible que no se obtenga 10 caras y 10 sellos; pero si lanzamos la misma moneda 2000 veces, es posible obtener un valor cercano a 1000 caras o 1000 sellos.

Si un experimento aleatorio tiene n resultados posibles y si A es un evento, de los cuales m son “favorables”, entonces la probabilidad P de que ocurra un suceso A estará dada por la relación $\frac{m}{n}$, es decir:

$$P(A) = \frac{m}{n}$$

En otras palabras:

$$P(A) = \frac{\text{Número de resultados favorables a } A}{\text{Total de resultados posibles del experimento}}$$

El total de resultados posibles del experimento no es otra cosa que el número de elementos (n) del espacio muestral, mientras que el número de resultados favorables a A es el número de elementos (m) del conjunto que describe al evento A .

Ejemplo:

En el experimento “lanzar un dado”, se pide determinar la probabilidad de obtener un resultado par.

Solución

Se determina el espacio muestral S el cual tiene 6 posibles resultados ($n = 6$) que son:

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

“ A ” representa el evento de obtener un resultado par, es decir, que A es un conjunto formado por tres resultados posibles, $A = \{2, 4, 6\}$ luego $m = 3$; entonces la probabilidad de obtener un resultado par es:

$$P(A) = \frac{3}{6} = 0,5$$

Luego existe un 50% de probabilidad que al lanzar un dado el resultado obtenido sea par.

Ejemplo:

En el lanzamiento de dos dados por ejemplo uno primero y luego el otro el espacio muestral es:

$$S = \{ (1,1) (1,2) (1,3) (1,4) (1,5) (1,6) (2,1) (2,2) (2,3) \\ (2,4) (2,5) (2,6) (3,1) (3,2) (3,3) (3,4) (3,5) (3,6) \\ (4,1) (4,2) (4,3) (4,4) (4,5) (4,6) (5,1) (5,2) (5,3) \\ (5,4) (5,5) (5,6) (6,1) (6,2) (6,3) (6,4) (6,5) (6,6) \}$$



$$n = 36$$

Determinar la probabilidad de obtener:

- A. En un dado 4 y en el otro 6 (sin importar el orden).
- B. En el primer dado 2 o 4 y en el segundo 3 o 5.
- C. 7 en la suma de los puntos.
- D. Un número mayor que 10 en la suma de los puntos.
- E. Resultados iguales en los dos dados.

Solución

- A. **R** representa el evento de obtener en un dado 4 y en el otro 6: $R = \{(4,6) (6,4)\}$, con $m = 2$

Luego, la probabilidad de obtener en uno de ellos el 4 y en el otro el 6 es:

$$P = \frac{2}{36} = \frac{1}{18} = 0,0555 \text{ o el } 5,55\%$$

- B. Sea **Q** el evento que representa el obtener en el primer dado 2 o 4 y en el segundo 3 o 5: $Q = \{(2,3)(2,5)(4,3)(4,5)\}$, entonces $m = 4$.

Luego:

$$P = \frac{4}{36} = 0,111 \text{ o el } 11,11\%$$

C. Sea **T** el evento de obtener 7 en la suma de los puntos: $T = \{(1,6) (2,5) (3,4) (6,1) (5,2) (4,3)\}$, con $m = 6$.

Luego, la probabilidad de obtener 7 en la suma de sus puntos es:

$$P = \frac{6}{36} = \frac{1}{6} = 0,1666 \text{ o el } 16,66\%$$

D. Sea **W** el evento obtener un número mayor que 10 en la suma de los puntos: $W = \{(5,6) (6,6) (6,5)\}$, donde $m = 3$.

Luego:

$$P = \frac{3}{36} = \frac{1}{12} = 0,0833 \text{ o el } 8,33\%$$

E. Sea **H** el evento obtener resultados iguales en los dados: $H = \{(1,1) (2,2) (3,3) (4,4) (5,5) (6,6)\}$, con $m = 6$.

Luego, la probabilidad de obtener resultados iguales en los dos dados es:

$$P = \frac{6}{36} = \frac{1}{6} = 0,1666 \text{ o el } 16,66\%$$

Como los eventos son conjuntos entonces entre ellos se pueden realizar operaciones tales como unión, intersección, y diferencia.

Luego se hablarán de las reglas que relacionan las operaciones entre eventos y la probabilidad.

4.1.1. Operaciones entre eventos y probabilidad

Ejemplo:

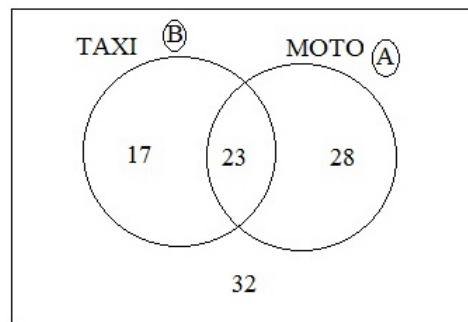
En un estudio en el colegio “Los Angeles” realizado a 100 estudiantes se les preguntó por el medio de transporte que utilizaban para ir al colegio.

Los resultados fueron:

* 23 utilizan moto y taxi.

- * 28 utilizan moto unicamente.
- * 17 utilizan unicamente taxi.
- * 32 llegan a pie.

Representando los resultados en un diagrama de Venn se tiene:



Entonces se pide calcular la probabilidad que al escoger un estudiante al azar, este

- ☆ Llegue al colegio utilizando moto y taxi.

Luego, el evento será $A \cap B$ y el número de elementos es $m = 23$ y hay un total de 100 estudiantes (o $n = 100$).

$$P(A \cap B) = \frac{23}{100} = 0,23$$

El 23% de los estudiantes utilizan como medio de transporte la moto y el taxi.

- ☆ Llegue al colegio utilizando moto o taxi.

Luego, el evento será $A \cup B$ y su probabilidad será:

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

$$P(A) = \frac{40}{100}$$

$$P(B) = \frac{51}{100}$$

$$P(A \cap B) = \frac{23}{100}$$

$$\begin{aligned}
 P(A \cup B) &= \frac{40}{100} + \frac{51}{100} - \frac{23}{100} \\
 &= \frac{68}{100} = 0,68
 \end{aligned}$$

El 68 % de los estudiantes utiliza moto o taxi para llegar al colegio.

☆ No utilice como medio de transporte el taxi.

Como A representa el evento utilizar taxi, entonces A^c representa el evento no utilizar taxi como medio de transporte.

$$\begin{aligned}
 P(A^c) &= 1 - P(A) \\
 &= 1 - \frac{40}{100} \\
 &= \frac{60}{100} \\
 &= 0,60
 \end{aligned}$$

El 60 % de los estudiantes no utiliza el taxi como medio de transporte.

Ejemplo:

De una baraja de 52 cartas, se extrae una de ellas. Hallar la probabilidad de que esa carta:

- Sea una Q
- no sea una Q.



Solución:

- Se define el suceso H como “La carta extraída es una Q”. Como la baraja tiene 4 cartas con Q (casos favorables), de las 52 cartas (casos posibles), entonces:

$$P(Q) = \frac{\text{Casos favorables}}{\text{Casos posibles}} = \frac{4}{52} = \frac{1}{13} = 0,0769 \text{ o } 7,6\%$$

Luego, existe el 7,6% de probabilidad de extraer una “Q” dentro de una baraja.

b. Para hallar la probabilidad del suceso: “La carta extraída no es una Q”.

Veamos: El suceso definido aquí es la negación del suceso del literal a); por lo tanto “La carta extraída no es una Q”, tiene como probabilidad:

$$P(\bar{Q}) = 1 - \frac{1}{13} = \frac{13 - 1}{13} = \frac{12}{13} = 0,923$$

Luego, la probabilidad de que la carta extraída no sea una Q es de 92,3%.

4.2. Probabilidad condicionada

En el cálculo de las probabilidades de algunos sucesos, el valor de dicha probabilidad varía en función del conocimiento de determinadas informaciones relativas a estos sucesos.

Ejemplo:

Se dispone de una urna que contiene cuatro bolas numeradas del 1 al 4, se extrae una bola y seguidamente se vuelve a introducir para realizar una segunda extracción, la probabilidad de extraer, por ejemplo, la bola número 3 en la segunda extracción es la misma que en la primera.

Si se realiza el mismo proceso sin reemplazar la bola extraída la probabilidad de extraer, por ejemplo, la bola número 3 en la segunda extracción dependerá de la bola extraída primero.

Probabilidad condicionada es la probabilidad de que ocurra un evento A, sabiendo que ha ocurrido otro evento B. La probabilidad condicional se escribe $P(A/B)$, y se lee “la probabilidad de A dado que B ya ocurrió”.

El conocimiento de que ha ocurrido el suceso B modifica, en algunas ocasiones, la probabilidad del suceso A, pero en otras no. Los sucesos en los que, conociendo que uno ha ocurrido,

no se modifica la probabilidad del otro, se dice que son independientes y, si se modifica, entonces son dependientes entre sí.

Ejemplo:

¿Cuál es la probabilidad de que al extraer tres cartas con reposición, de una baraja de 40, estas sean “as”, “rey de copas ” y “seis ”?.

Cuando se dice con *reposición* se entiende que al extraer la primera carta para ser observada, se regresa nuevamente al *mazo*, es decir, la baraja queda completa con sus cuarenta cartas, procediendo después a una nueva extracción.

**Solución**

Sean los sucesos: A (ases), B (rey de copas) y C (cartas con seis puntos).

Se tienen 4 “ases” $P(A) = \frac{4}{40}$

Se tienen un “rey de copas” $P(B) = \frac{1}{40}$

Hay 4 “ seis ” $P(C) = \frac{4}{40}$

Se observa que los resultados son independientes, pues ninguno de ellos se ve afectado por la ocurrencia del otro; en este caso se aplica la regla de multiplicación, siendo:

$$\begin{aligned} P(A \cap B \cap C) &= P_{(A)} \times P_{(B)} \times P_{(C)} \\ &= \frac{4}{40} \times \frac{1}{40} \times \frac{4}{40} \\ &= \frac{16}{64000} \\ &= \frac{1}{4000} \\ &= 0,00025 \text{ o } 0,025 \% \end{aligned}$$

Luego, al extraer tres cartas con reposición de un mazo de 40 la probabilidad de que estas cartas sean “as”, “rey de copas ” y “seis ” es de 0,025 %.

4.3. Técnicas de conteo

Son métodos que sirven para determinar el número de elementos de un espacio muestral o de un evento teniendo en cuenta dos características importantes el orden y/o la repetición. Además no se hace necesario conocer cada elemento.

Entre las formas de conteo se tienen: el principio de multiplicación, permutaciones, y combinaciones.

4.3.1. Principio de multiplicación

Ejemplo:

Karla cuenta con 3 blusas y 2 pantalones se quiere determinar de cuantas formas distintas se puede vestir Karla?

$$3 \text{ blusas} \times 2 \text{ pantalones} = 6 \text{ maneras diferentes}$$

Ejemplo:

Julian va al restaurante a almorzar y encuentra que hay 3 sopas diferentes, 4 secos y 2 postres ¿Cuántas formas posibles tiene Julian para pedir su almuerzo?

$$3 \text{ sopas} \times 4 \text{ secos} \times 2 \text{ postres} = 24 \text{ almuerzos diferentes}$$

Si un experimento está formado por n - ensayos realizados en un orden específico en donde el primero tiene N_1 resultados posibles, el segundo N_2 y así sucesivamente. Entonces el número de resultados posibles para el experimento serán: $N_1 \times N_2 \times \dots \times N_n$

Explicación, en el ejemplo 1 el experimento está conformado por dos ensayos: escoger Blusa y escoger Pantalón. El primer ensayo tiene 3 posibles resultados ($N_1 = 3$) y el segundo tiene 2 posibles resultados ($N_2 = 2$). Entonces el número de resultados posibles para el experimento es de 6 ($N_1 \times N_2 = 3 \times 2 = 6$)

Ejemplo:

Un uniforme deportivo se fabrica en cinco estilos diferentes y en cuatro colores distintos para cada uno. ¿De cuántas formas distintas se puede organizar el uniforme deportivo?

Solución

$$5 \text{ estilos} \times 4 \text{ colores} = 20 \text{ uniformes diferentes}$$

Ejemplo:

En una marquería los tres últimos pasos para terminar la enmarcación de un cuadro consiste en: lijar el marco, pintar el marco y, por último colocar la foto en el marco. Para lijar el marco hay disponibles cuatro operarios, para pintar el marco hay tres operarios y para colocar la foto dos operarios. En cada paso el operario correspondiente se encarga específicamente de dicho cuadro, según esto, ¿de cuántas formas se puede terminar la enmarcación de un cuadro?

Solución

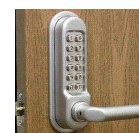
$$4 \text{ lijar el cuadro} \times 3 \text{ pintar} \times 2 \text{ colocar la foto} = 24 \text{ Maneras de enmarcar un cuadro.}$$

Normalmente usamos la palabra “combinación” descuidadamente, sin pensar en si el orden de las cosas es importante. En otras palabras:

- “Mi ensalada de frutas es una combinación de manzanas, uvas y bananas”: no importa en qué orden van las frutas, podría ser “bananas, uvas y manzanas” o “uvas, manzanas y bananas”, es la misma ensalada.



- “La combinación de la cerradura es 472”: ahora sí importa el orden. “724” no funcionaría, ni “247”. Tiene que ser exactamente $4 - 7 - 2$.



En matemáticas se utiliza un lenguaje más preciso:

- ✎ Si el orden no importa, es una **combinación**.
- ✎ Si el orden sí importa es una **permutación**.

Veamos:

4.3.2. Permutaciones

Se denominan *permutaciones* de n elementos, los diferentes grupos que se pueden hacer tomándolos todos de una vez. Las permutaciones implican orden en la ubicación de los elementos.

Se pueden hacer permutaciones sin sustitución, con sustitución y permutaciones r sin repetición.

Permutaciones con sustitución

Cuando se cuenta la cantidad de formas como se puede ordenar un conjunto de objetos, donde cada objeto puede aparecer más de una vez, se está contando las **permutaciones con sustitución del conjunto**

El número de permutaciones con sustitución de un conjunto de n elementos se calcula así:

$$\underbrace{n \times n \times n \times n \times \dots \times n \times n}_{m \text{ veces}} = n^m$$

Permutaciones sin sustitución

Cuando se cuenta la cantidad de formas como se puede ordenar un conjunto de objetos, donde cada objeto debe aparecer una sola vez, se está contando las **permutaciones sin sustitución del conjunto**.

El número de permutaciones sin sustitución de un conjunto de n elementos se calcula de la siguiente forma:

$$P_n = n \times (n - 1) \times (n - 2) \times (n - 3) \times \dots \times 3 \times 2 \times 1 = n!$$

donde n representa la cantidad de elementos del conjunto.

En este caso, se reduce el número de opciones en cada paso.

Ejemplo:

Juan, Camilo, Paola, Homero y Lisa están esperando la ruta para dirigirse al trabajo en el mismo paradero. ¿De cuántas formas distintas pueden subir al bus?

Solución

En este experimento se considera el orden en el que suben al bus y no existe la repetición ya que un trabajador no puede subir dos veces. La cantidad de elementos del conjunto es cinco ($n = 5$). Luego el número de elementos del espacio muestral será:

$$P_n = n! = 5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$$

Los cinco trabajadores pueden subir al bus de 120 formas distintas.

Ejemplo:

¿De cuántas formas distintas se podrían ordenar 16 bolas de billar?

Solución Como se quieren ordenar las 16 bolas que conforman el juego, entonces para escoger la primera bola hay 16 opciones, para escoger la segunda hay 15 opciones, para escoger la tercera hay 14 opciones y así sucesivamente 13, 12, 11, 10, 9, 8, ..., luego se tiene:



$$16 \times 15 \times 14 \times 13 \dots \times 2 \times 1 = 20922789888000 \text{ opciones}$$

Entonces existen 20922789888000 formas distintas de ordenar las 16 bolas de billar.

Permutaciones r sin repetición

Ejemplo:

María tiene 5 rutas distintas para ir al colegio, pero en un día solo puede utilizar 2. ¿De cuántas maneras puede María escoger la ruta para ir y venir del colegio?

Solución

La forma de contar las posibilidades, consiste en suponer que para ir al colegio tiene 5 posibilidades y para el regreso, como ya utilizó una y no desea repetirla, entonces solo tiene cuatro posibilidades. En total tiene 20 posibilidades para ir y regresar del colegio. Es decir:

Permutaciones de 2 objetos de 5: $5 \times 4 = 20$

Una ordenación de solamente r elementos de un total de n objetos disponibles se llama una permutación r .

El número de permutaciones de r objetos de un total de n sin repetición, se calcula así:

$${}_nP_r = n \times (n - 1) \times (n - 2) \times (n - 3) \times \dots \times (n - r + 1) = \frac{n!}{(n - r)!}$$

Ejemplo:

De un total de 6 personas se desean escoger 3 para que representen al grupo en un evento en la ciudad de Cali. ¿Cuántas formas hay de escogerlas?



Solución

La primera persona se puede escoger entre las seis posibilidades; para la segunda persona solo quedan 5 posibilidades; para la tercera y última persona se puede escoger entre las 4 restantes.

En total son:

$P(6, 3) = 6 \times 5 \times 4 = 120$ formas de escoger a los representantes. o

$$\begin{aligned} {}_6P_3 &= \frac{6!}{(6 - 3)!} \\ &= \frac{6!}{3!} \\ &= \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1} \\ &= 120 \end{aligned}$$

Ejemplo:

Cuatro equipos disputan un campeonato que clasifica a solo dos equipos. El primer clasificado irá al mundial de la categoría y el segundo clasificado irá a un torneo europeo. ¿De cuántas formas distintas pueden dos de los cuatro equipos clasificar al mundial y al torneo?

Solución

Existe orden ya que el primer clasificado tiene un premio distinto al segundo. No hay repetición ya que un equipo no puede asistir a ambos torneos. El número de elementos del espacio muestral es:

$$\begin{aligned} {}_4P_2 &= \frac{4!}{(4-2)!} \\ &= \frac{4!}{2!} \\ &= \frac{4 \times 3 \times 2 \times 1}{2 \times 1} \\ &= 12 \end{aligned}$$

Existen 12 formas distintas de lograr que dos equipos de cuatro asistan al torneo.

4.3.3. Combinaciones

Es la forma de organizar los elementos de un conjunto sin tener en cuenta su orden.

El método para saber el número de estas combinaciones consiste encontrarlas como si fueran permutaciones r , es decir, $\frac{n!}{(n-r)!}$, y como en las combinaciones no se tiene en cuenta el orden, entonces se divide entre el número de veces que se repite cada permutación, así:

$${}_nC_r = P(n, r) \frac{1}{r!}$$

$${}_nC_r = \frac{n!}{(n-r)!} \times \frac{1}{r!}$$

$$\boxed{{}_nC_r = \frac{n!}{(n-r)! r!}}$$

Ejemplo:

Hallar el número de formas en que se pueden mezclar cuatro colores: rojo (r), verde (v), azul (a) y blanco (b), tomándolos de tres en tres.

Solución

El experimento consiste en mezclar los 4 colores tomándolos de a 3.

En este caso, el orden no se tiene en cuenta ya que al combinar el color rojo, el verde y el blanco se obtiene el mismo resultado que si se toma primero el verde, luego el blanco y por último el rojo.

El número de formas posibles es:

$${}^4C_3 = \frac{4!}{(4-3)!3!} = \frac{4 \times 3 \times 2 \times 1}{1! \times 3!} = \frac{4 \times 3 \times 2 \times 1}{3 \times 2 \times 1} = 4$$

Por tanto, el número de formas es 4. Ellos son:

$$\{(r, v, a), (r, v, b), (r, b, a), (b, a, v)\}$$

Ejemplo:

¿Cuántos comités de 3 miembros se pueden seleccionar de un grupo de 8 personas?

Solución

Consiste en hacer grupos o comités de 3 personas dentro de un total de 8 personas, en este experimento el orden no se tiene en cuenta por qué no importa a quien se elija de primero, a quien de segundo y a quien de tercero, solamente importa que en cada comité se encuentren 3 personas. ($n = 8$ y $r = 3$)

$${}^8C_3 = \frac{8!}{(8-3)!3!} = \frac{8!}{5! \times 3!} = \frac{8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{5 \times 4 \times 3 \times 2 \times 1 \times 3 \times 2 \times 1} = \frac{8 \times 7 \times 6}{3 \times 2 \times 1} = 56$$

Luego, se pueden formar 56 comites diferentes.

Ejemplo:

Se le pide a una persona que seleccione al azar cuatro cartas de una baraja de 52. ¿De cuántas formas distintas se puede hacer esto?

Solución

No importa el orden ya que las cartas representan lo mismo sin importar cual salió primero

.

El número de formas de escoger las cuatro cartas son:

$${}^4C_{52} = \frac{52!}{(52-4)!4!} = \frac{52!}{48! \times 4!} = \frac{52 \times 51 \times 50 \times 49 \times 48!}{48! \times 4 \times 3 \times 2 \times 1} = \frac{6364800}{24} = 265200$$

Entonces existen 265200 formas de seleccionar cuatro cartas de una baraja de 52.

4.4. Talleres de aplicaciones

1. El Sr. Fernández está dudando entre gastar sus ahorros en un viaje a las Islas Galápagos o invertir en renta variable. Su asesor fiscal le ofrece dos alternativas atractivas, pero él ante su falta de formación bursátil, confía al azar su decisión. Invertirá en el sector eléctrico si saca una bola roja de una urna que contiene 20 bolas, de las cuales 8 son rojas, 3 verdes y 9 negras. Si la bola no es roja lanzará dos dados y si obtiene una suma de 6 entre ambos invertirá en el sector inmobiliario; en caso contrario se decidirá por las vacaciones en las Galápagos. ¿Cuál es la probabilidad de que finalmente disfrute del viaje?.



2. Una urna contiene 4 bolas blancas y 3 rojas. Se efectúan 2 extracciones sucesivas. Obtener la probabilidad de extraer una bola blanca y a continuación una bola roja:
- Quando habiendo extraído la primera bola se devuelve a la urna para realizar la segunda extracción.
 - Quando habiendo extraído la primera bola no se devuelve a la urna para realizar la segunda extracción.
3. Una urna contiene 7 papeletas numeradas del 1 al 7.
- Si se sacan 2 papeletas sin remplazamiento, ¿cuál es la probabilidad de que las dos, tengan números impares?
 - ¿Cuál es la probabilidad de que ocurra lo mismo siendo la extracción con remplazamiento?
4. De cuantas formas pueden ordenarse 6 libros en una biblioteca si:
- No se da ninguna restricción.
 - 3 libros determinados determinados deben estar juntos.
 - 1 libro determinado debe estar en el extremo derecho.
-

-
5. Calcular el número de permutaciones que se pueden formar con las letras de la palabra **Mejico**.
 6. En un grupo hay 20 mujeres y 15 hombres. Se escoge al azar un equipo de 5 personas. ¿Cuál es la probabilidad de que todas sean mujeres?
 7. El equipo de baloncesto del colegio “Aprender jugando” está conformado por 12 estudiantes. Conteste:
 - a. ¿Cuántas maneras hay de asignar las 5 posiciones de juego del equipo?
 - b. ¿Cuántas maneras hay de asignar las posiciones del equipo si una de ellas solo puede ser ocupada por Jaime Perez?
 8. En una presentación musical se quieren sentar 5 hombres y 4 mujeres en una fila. Señale de cuántas formas se pueden acomodar si:
 - a. Pueden sentarse en cualquier lugar.
 - b. Las mujeres y los hombres deben sentarse juntos.
 - c. Los hombres deben sentarse en los sitios impares.
 9. Un estudiante tiene un examen de 10 preguntas y debe contestar 7 de las 10.
 - a. ¿De cuántas maneras puede elegir las?
 - b. Si las cuatro primeras preguntas son obligatorias, ¿de cuántas formas puede elegir las preguntas?
 10. En un barrio hay 100 adolescentes, 40 de los niños y 35 de las niñas juegan fútbol. El total de niñas en el barrio es de 45. Si se elige un adolescente del barrio al azar, se pide determinar:
 - a. ¿La probabilidad de que sea niño?
 - b. Si sabemos que juega fútbol, ¿cuál es la probabilidad de que sea chica?
-

- c. ¿Cuál es la probabilidad de que sea un chico que no juegue fútbol?
11. En una universidad, el 30% de los alumnos hablan inglés, el 10% frances y el 8% los dos idiomas. ¿Cuál es la probabilidad de encontrar alumnos que hablen alguna lengua extranjera?
-

PRUEBAS SABER 11^o

PREGUNTAS DE SELECCIÓN MÚLTIPLE CON ÚNICA RESPUESTA - TIPO I

Estas preguntas, se desarrollan en torno a una idea o un problema y constan de un enunciado y cuatro opciones de respuesta. Se recomienda leer cuidadosamente el enunciado y, después de analizarlo, escoger entre las opciones la que se considera correcta.

1. La frecuencia relativa expresada en forma porcentual es igual a la frecuencia relativa:
 - A. Multiplicada por 100.
 - B. Divida por 10.
 - C. Multiplicada por 10.
 - D. Divida por 100.
2. Las frecuencias relativas suman siempre:
 - A. 1.
 - B. 100.
 - C. 0.
 - D. Cualquier número.

3. Una distribución de frecuencias puede tener:
- A. Más de una media aritmética.
 - B. Ninguna media aritmética.
 - C. Solo una media aritmética.
 - D. Ninguna de las anteriores.
4. Un experimento aleatorio se caracteriza por
- A. El experimento se puede repetir indefinidamente bajo las mismas condiciones
 - B. Cualquier modificación en las condiciones iniciales puede modificar el resultado final
 - C. Se puede determinar el conjunto de los posibles resultados del experimento, pero no se puede predecir previamente un resultado
 - D. Todas las anteriores son correctas
5. El conjunto de todos los sucesos posibles de un experimento aleatorio es
- A. Una unión de sucesos
 - B. Un suceso compuesto
 - C. El espacio muestral
 - D. Ninguna de las anteriores
6. La diferencia entre el dato mayor y el dato menor de una distribución recibe el nombre de:
- A. Moda.
 - B. Desviación.
 - C. Rango.
-

D. Mediana.

**RESPONDA LAS PREGUNTAS 7 Y 8 DE ACUERDO CON LA
SIGUIENTE INFORMACIÓN**

Cada una de las placas de las carrozas de un municipio contiene tres letras. La primera letra se escoge del conjunto C, B, L, W, R, la segunda letra se escoge del grupo E, I, U y la tercera del conjunto P, M, N, T.

7. ¿Cuál es la probabilidad que al tomar una carroza en este municipio esta tenga la placa CUT?
- A. $\frac{1}{3}$
 - B. $\frac{3}{60}$
 - C. $\frac{1}{60}$
 - D. $\frac{1}{12}$
8. Cuando hubo que expedir más placas en el municipio, el alcalde determinó añadir dos nuevas letras. Se puede añadir las dos letras nuevas a uno de los conjuntos, o bien, se puede añadir una letra nueva a uno de los conjuntos y la otra letra nueva a otro conjunto. ¿Cuál es el mayor número de nuevas placas que se pueden hacer cuando se añaden las dos nuevas letras?
- A. 24
 - B. 40
 - C. 36
 - D. 60
-

**RESPONDA LAS PREGUNTAS 9 Y 10 DE ACUERDO CON LA
SIGUIENTE INFORMACIÓN**

Una caja contiene exactamente cinco fichas, tres rojas y dos blancas. Se sacan fichas, una o la vez sin reemplazar, hasta que se hayan extraído o bien todas las fichas rojas o bien todas las fichas blancas.

9. ¿Cuál es la probabilidad de sacar primero una ficha roja?

- A. $\frac{1}{5}$
- B. $\frac{3}{5}$
- C. $\frac{1}{10}$
- D. $\frac{2}{10}$

10. ¿Cuál es la probabilidad de que queden en la urna las tres fichas rojas?

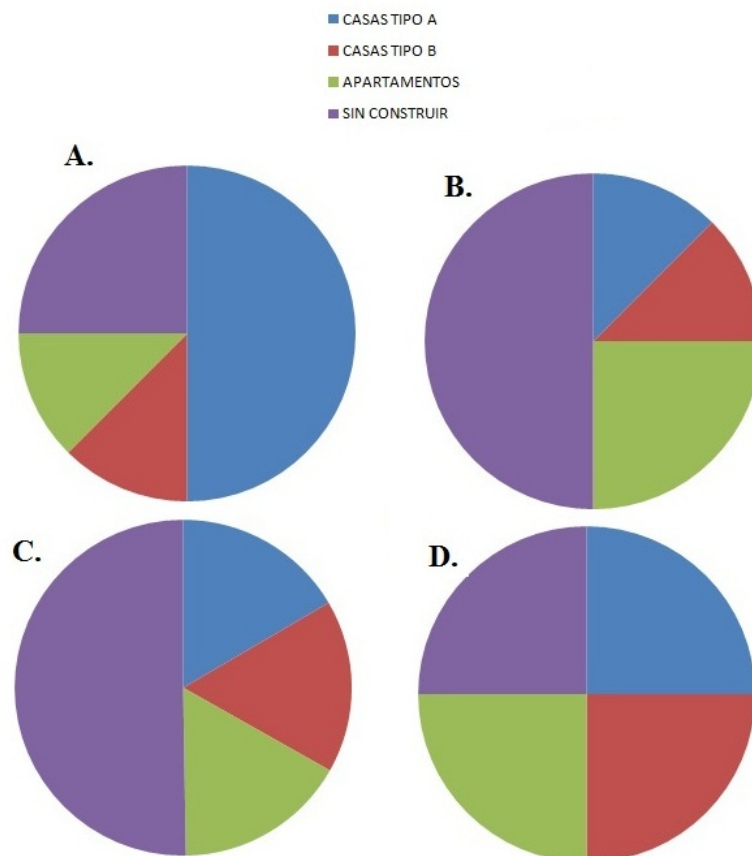
- A. $\frac{2}{5}$
- B. $\frac{3}{5}$
- C. $\frac{1}{10}$
- D. $\frac{3}{10}$

11. Un profesor de educación física encontró que el 40 % de sus estudiantes juegan fútbol, el 20 % baloncesto y un 25 % de los restantes practican atletismo. ¿Qué porcentaje de los estudiantes juegan un deporte distinto a los anteriores, sabiendo que ningún estudiante practica 2 deportes?

- A. El 15 % exactamente
 - B. Menos del 20 %
 - C. El 30 % exactamente
-

D. Más del 30 %

12. Una constructora desarrolla un proyecto de vivienda social. En la primera etapa se utiliza exactamente el 50 % de la totalidad del terreno destinado para la construcción de casas. En dicho terreno se construyó un 25 % de casas tipo A y en $\frac{1}{4}$ del mismo terreno casas tipo B; y en la parte restante apartamentos. El diagrama que muestra el terreno que fue asignado para construir apartamento es:



13. La siguiente tabla muestra algunos datos sobre el número de computadores por hogar obtenidos en una muestra de 200 hogares elegidos al azar, usando el directorio telefónico de la ciudad.

No. de computadores	0	1	2	3
Frecuencia Absoluta		120		
Porcentaje	5 %		27 %	

- ¿Cuál de las siguientes afirmaciones no corresponde a la información que se obtiene en la tabla?
- A. 54 hogares tienen 2 computadores
 - B. El 10 % de los hogares tiene 3 computadores
 - C. El 60 % de los hogares tiene un computador
 - D. 10 hogares no tienen computador
14. Se realizará un sondeo sobre la gestión de cierto funcionario dentro del gobierno nacional. Para tal propósito se seleccionarán algunos colombianos mayores de edad para que den su opinión sobre el tema. Si se busca que los resultados sean lo más objetivos posibles, ¿cuál de los siguientes grupos debería ser el entrevistado?
- A. 1000 personas de cada estrato social de una ciudad.
 - B. 6000 hombres de cualquier estrato social de una ciudad.
 - C. 1000 personas de cada estrato social de diferentes ciudades.
 - D. 6000 mujeres de cualquier estrato social de diferentes ciudades.
-

15. La historia clínica de 120 pacientes de un hospital se resume en el siguiente cuadro:

	Fuma	No fuma	Total
Afección cardiaca	45	25	70
Otra enfermedad	20	30	50
Total	65	55	120

¿Cuál es la probabilidad que al elegir un paciente que fuma, este no tenga una afección cardiaca?

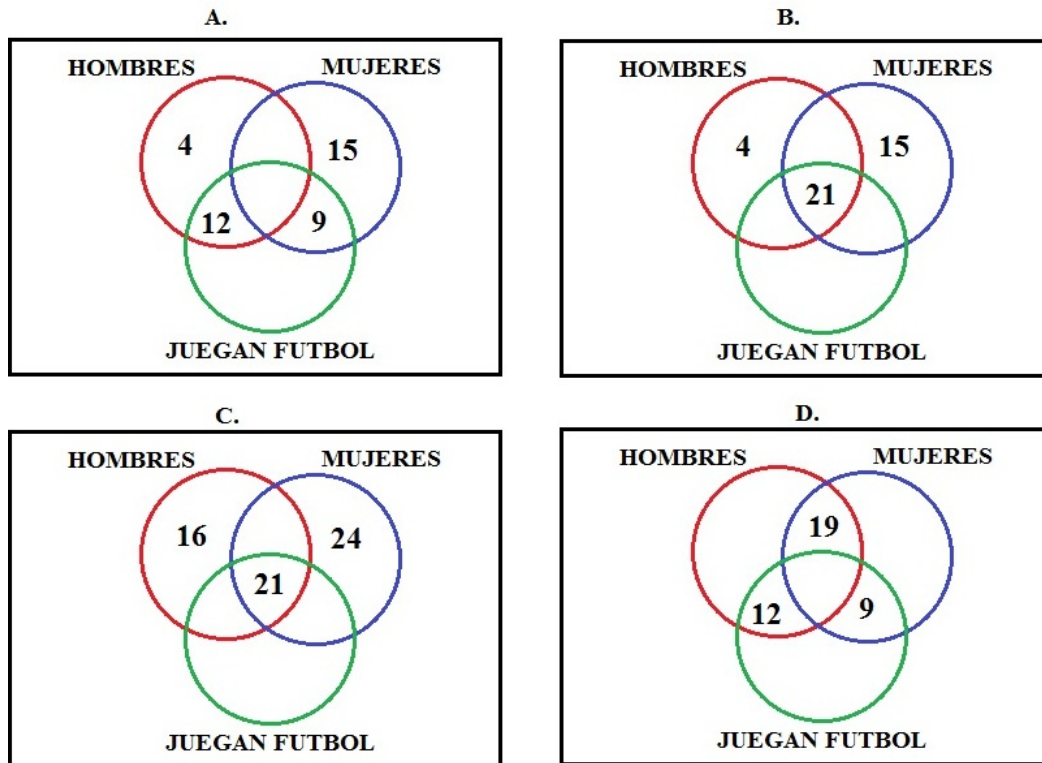
- A. $\frac{45}{65}$
- B. $\frac{20}{65}$
- C. $\frac{20}{120}$
- D. $\frac{65}{120}$

RESPONDA LAS PREGUNTAS 16 A 18 DE ACUERDO CON LA SIGUIENTE INFORMACIÓN

En el salón del grado décimo donde hay 40 estudiantes se realizó una encuesta sobre quienes practican el fútbol. La tabla representa la información recogida

	Juegan fútbol	No juegan futbol
Hombres	12	4
Mujeres	9	15

16. ¿Cuál de las siguientes gráficas representa la información registrada en la anterior tabla?



17. El colegio tiene programado realizar una escuela de fútbol. ¿Qué porcentaje aproximado del salón podría estar interesado?

- A. 12 %
 B. 21 %
 C. 40 %
 D. 52 %

18. Se selecciona un estudiante al azar, la probabilidad que el estudiante sea mujer y juegue fútbol es de:

- A. $\frac{9}{15}$

- B. $\frac{9}{21}$
- C. $\frac{9}{24}$
- D. $\frac{9}{40}$

**RESPONDA LAS PREGUNTAS 19 Y 20 DE ACUERDO CON LA
SIGUIENTE INFORMACIÓN**

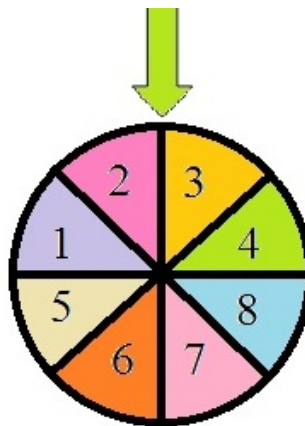
Andres y Camilo están jugando con tres dados corrientes a ver quien acierta primero, sin importar el orden, los seis puntajes que escribieron en sus respectivos cuadernos. Cada jugador, en su turno, lanza los tres dados y suma los puntos visibles de las caras superiores. Andres escribió (7, 8, 9, 10, 11, 12); mientras que Camilo escogió para jugar (3, 4, 5, 16, 17, 18)

19. De los dos quien puede ganar con mayor facilidad es:
- A. Camilo porque sus puntajes tienen mayor probabilidad.
 - B. Cualquiera por que ganar depende de la suerte.
 - C. Andres porque los puntajes de Camilo tienen menor probabilidad.
 - D. Cualquiera por que el espacio muestral es equiprobable
20. El juego ha transcurrido, a Andres le queda por obtener 10 y 12, mientras que ha camilo le quedan el 3 y el 18. De esta situación se puede deducir que:
- A. Ambos tienen las mismas probabilidades de ganar porque tienen dos puntajes.
 - B. Es más fácil que gane Camilo por que sus puntajes son extremos.
 - C. Es mas fácil que gane Andres porque sus puntajes son medios.
 - D. Cualquiera pude ganar porque ambos tienen las mismas probabilidades en los números que les queda.
-

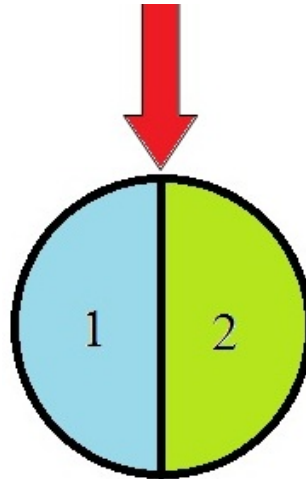
21. En una bolsa negra se introducen dos balotas blancas, tres rojas y cinco azules. Se extraen de la bolsa tres balotas azules de manera consecutiva, las cuales no se vuelven a introducir, por lo que alguna persona afirma que es más fácil que la cuarta balota que se extraiga sea también de color azul. Esta afirmación es:
- A. Falsa porque este hecho hace que la probabilidad de sacar una balota roja sea igual que sacar una balota azul.
 - B. Falsa porque lo que se logra con esto es que se haga mayor la probabilidad de sacar una balota roja por encima de las demás.
 - C. Verdadera porque hay todavía más balotas azules.
 - D. Verdadera porque entre más boletas azules se extraigan, mayor se vuelve la probabilidad de sacar una balota azul.

RESPONDA LAS PREGUNTAS 22 A 24 DE ACUERDO CON LA SIGUIENTE INFORMACIÓN

Juanita y Adriana comparan las posibilidades que hay al jugar en sus ruletas. Juanita tiene la siguiente ruleta:



Y Adriana esta



El juego de Juanita consiste en adivinar cuál es el número que va a señalar la felcha y el de Adriana consiste en adivinar la terna (es decir, girar la ruleta tres veces y descubrir que números y en que orden van a salir).

22. Hay mas posibilades de ganar en el juego de:
- A. Juanita
 - B. Adriana
 - C. Cualquiera de los dos
 - D. No es posible saberlo
23. Natalia quiere elaborar una ruleta donde la probabilidad de ganar sea mayor que la de Adriana. Para ello bastaría con:
- A. Hacer una ruleta igual a la de Juanita.
 - B. Elaborar una ruleta con tres números en renglones de igual área, formando parejas.
 - C. Construir una ruleta igual a la de Adriana y formar parejas.
 - D. Hacer una ruleta igual a la de Juanita y formar parejas.
-

24. Juanita ha decidido ahora jugar girando la ruleta tres veces. ¿Cuál de los siguientes resultados es más probable que ocurra?
- A. Obtener 2, 1, 5, en este orden exactamente.
 - B. Obtener 2, 1, 5 en cualquier orden.
 - C. Obtener 1, 1, 5 en cualquier orden
 - D. Obtener 1, 1, 1.

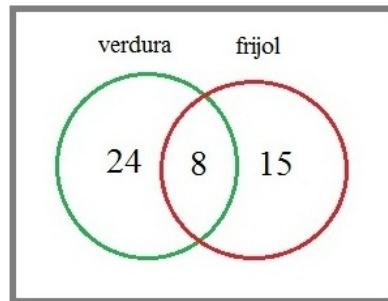
**RESPONDA LAS PREGUNTAS 25 Y 26 DE ACUERDO CON LA
SIGUIENTE INFORMACIÓN**

- Hay dos jugadores, a cada jugador se le entregan tres fichas: la ficha 1 tiene una “A” en una cara y una “B” en la otra; la ficha 2 una “A” en una cara y una “C” en la otra y la tercera ficha con una “B” en una cara y una “C” en la otra.
25. Uno de los jugadores tira las tres fichas a la vez encima de una mesa. Gana el jugador 1 si coinciden dos fechas cualesquieras; gana el jugador 2 si salen las tres diferentes. De acuerdo con la información anterior se puede decir que.
- A. El jugador 1 no tiene oportunidad de ganar.
 - B. El jugador 2 tiene más del 60 % de probabilidades de ganar.
 - C. El jugador 1 tiene mayor probabilidad de ganar que el jugador 1.
 - D. Los dos jugadores tienen la misma posibilidad de ganar.
26. ¿Cuál de las siguientes sucesiones es más probable que resulte al lanzar la ficha 1 cinco veces consecutivas?
- A. AAABB
 - B. ABAAA
 - C. BABAB
-

- D. Cualquiera de las anteriores
27. En el barrio Reinaldo, la Registraduría Nacional dispone una mesa para que los ciudadanos se inscriban con el fin de participar en las próximas elecciones.
- Teniendo en cuenta las probabilidades es más cercano a lo que pueda pasar afirmar que:
- A. De las próximas 10 inscripciones, 8 corresponda a hombres.
 - B. De las próximas 100 inscripciones 80 correspondan a hombres.
 - C. Cualquiera sea el número de inscritos la mitad será hombre y la mitad mujeres.
 - D. A y B son igualmente probables.
28. En un concurso mundial de matemáticas participan 5 americanos, 7 europeos, 3 asiáticos, 2 africanos y 1 australiano. ¿Cuál es la probabilidad de que el ganador del concurso mundial de matemáticas no sea americano?
- A. $\frac{1}{18}$
 - B. $\frac{5}{18}$
 - C. $\frac{13}{18}$
 - D. $\frac{17}{18}$

**RESPONDA LAS PREGUNTAS 29 Y 30 DE ACUERDO CON LA
SIGUIENTE INFORMACIÓN**

El siguiente diagrama representa la elección de los clientes en un restaurante, respecto a las opciones de principio ofrecidas en el menú en el día: frijol o verdura.



29. ¿Cuántas personas no incluyen el frijol como principio ?
- A. 8
 - B. 15
 - C. 23
 - D. 24
30. ¿Cuál de las siguientes afirmaciones es verdadera?
- A. Solamente 15 personas escogieron frijol
 - B. 39 personas pidieron el menú del día
 - C. 8 personas pidieron frijol y verduras
 - D. No se puede saber exactamente cuántos escogieron ambas opciones
31. Dentro de un grupo de personas encuestadas se encontró que el 45 % está conforme con las opciones de programas de televisión que se presentan en la noche, un 43 % no está conforme con estas opciones y las 48 personas restantes no saben o no responden.
- La cantidad de personas que fueron encuestadas es:
- A. 100
 - B. 200
 - C. 300
-

D. 400

32. De una urna que contiene y balotas blancas y t balotas amarillas, se van extrayendo balotas que se introducen nuevamente, ¿Cuál es la probabilidad de que en la n ésima extracción se obtenga una balota blanca?

A. $\frac{y}{t}$

B. $\frac{t}{y}$

C. $\frac{t}{y+t}$

D. $\frac{y}{y+t}$

RESPONDA LAS PREGUNTAS 33 Y 34 DE ACUERDO CON LA SIGUIENTE INFORMACIÓN

33. Se hizo una encuesta de consumo, dentro de la cual se preguntaba cuántos vasos de agua se tomaba cada uno diariamente. La siguiente presenta los intervalos presentados en las opciones y las frecuencias acumuladas resultantes

Intervalo de consumo	Frecuencia acumulada
Más de 0 y menos de 3	33
Más de 3 y menos de 6	74
Más de 6 y menos de 9	152
Más de 9 y menos de 12	167
Más de 12	170

¿Cuál es la moda de los resultados obtenidos?

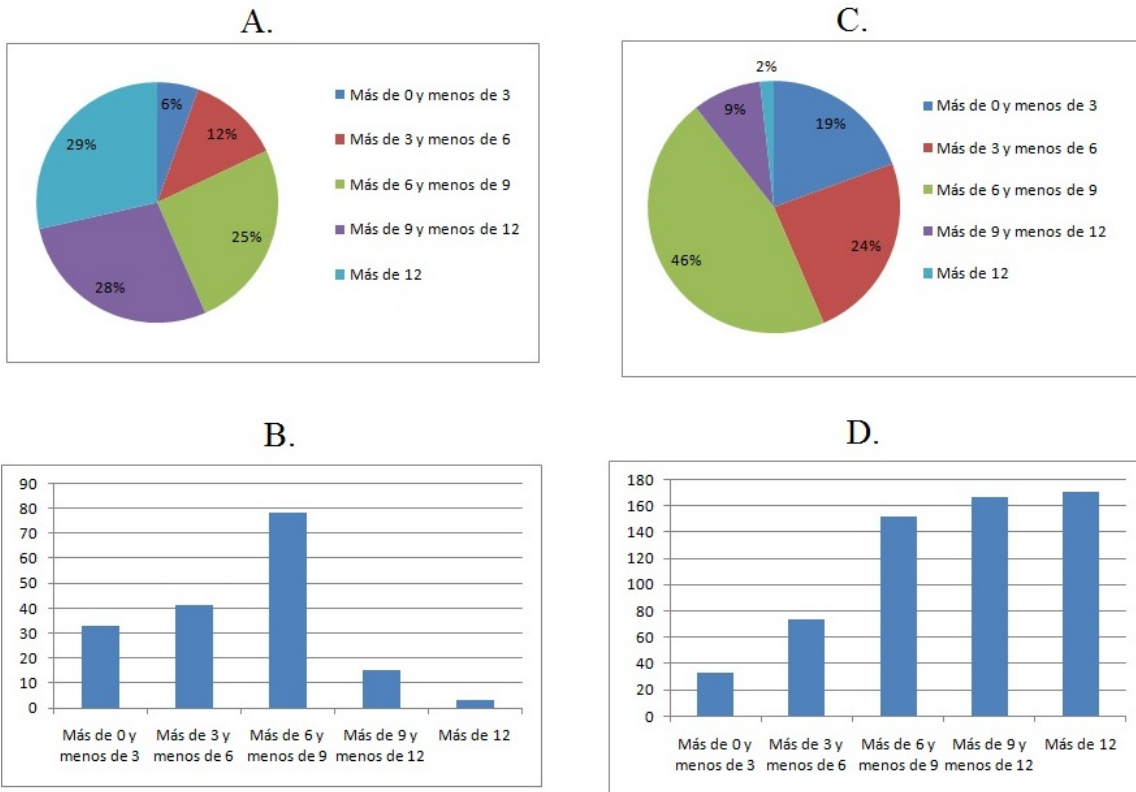
A. Más de 3 y menos de 6

B. Más de 6 y menos de 9

C. Más de 9 y menos de 12

D. Más de 12

34. ¿Cuál de las siguientes gráficas representa los porcentajes de los resultados obtenidos?



RESPONDA LAS PREGUNTAS 35 Y 36 DE ACUERDO CON LA SIGUIENTE INFORMACIÓN

En la fábrica CHOCOL S.A se ha hecho un avalúo de las ventas, los costos de producción y mano de obra de sus productos. También se han distribuido los valores de las ganancias según el tipo de producto (de cocina, dulcería, granulados y heladería) que se hayan vendido.

Resultó entonces lo siguiente:

Concepto	Cantidad de dinero
Ventas	20 millones de dólares
Costos de producción	9 millones de dólares
Mano de obra, transporte y administrativo	7 millones de dolares

Ganancias optenidas por vender

De cocina	2 millones de dólares
Dulceria	1 millon de dólares
Heladeria	600000 dolares
Granulados	400000 dolares

35. ¿Cuál o cuáles de las siguientes afirmaciones es ó son verdadera(s)?

- I. El 5 % de las ganancias se obtuvieron por vender productos de dulcería
- II. El 10 % de las ganancias se obtuvieron por vender productos de cocina
- III. El 15 % de las ganancias se obtuvieron por vender productos de heladería

- A. I solamente
- B. III solamente
- C. I y II solamente
- D. II y III solamente

36. Respecto a las ganancias obtenidas por vender es correcto afirmar que:

- A. No alcanzan a cubrir la inversión de costo de producción y mano de obra, transporte y administrativos.

- B. Sobrepasan la diferencia entre las ventas y los costos de producción
- C. Se obtienen al calcular la diferencia entre ventas y la suma de costo de producción y mano de obra, transporte y administrativos
- D. Se obtienen al calcular la diferencia entre costo de producción y mano de obra, transportes y administrativos.

**RESPONDA LAS PREGUNTAS 37 Y 38 DE ACUERDO CON LA
SIGUIENTE INFORMACIÓN**

Una rifa consta de boletas numeradas desde el número 000 y el 999 distribuidas en 20 talonarios.

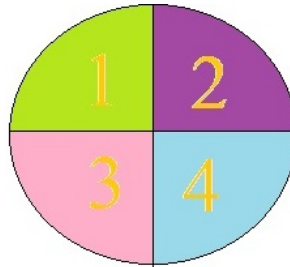
37. Juan está vendiendo uno de esos talonarios , ¿Cuál es la probabilidad de que la boleta ganadora de la rifa sea del talonario de Juan?

- A. $\frac{1}{20}$
- B. $\frac{1}{50}$
- C. $\frac{20}{50}$
- D. $\frac{30}{50}$

38. Si se inventa un plan de premios en los que las boletas que tengan las mismas cifras del número ganador ¿Cuál es la máxima probabilidad de sacar una boleta premiada?

- A. $\frac{1}{1000}$
 - B. $\frac{1}{500}$
 - C. $\frac{3}{1000}$
 - D. $\frac{3}{500}$
-

39. Un tablero de tiro al blanco distribuye sus puntajes de tiro al blanco de la siguiente manera:



- Cada jugador tiene derecho a lanzar dos dardos, y en caso de que el dardo caiga por fuera del tablero el puntaje correspondiente al lanzamiento es cero. Si en el segundo lanzamiento da en el tablero en la misma región del primer lanzamiento, se repite hasta que caiga en una distinta. ¿Cuáles son todos los posibles resultados que puede obtener un jugador terminado su turno?
- A. 1, 2, 3, 4
B. 5, 6, 7, 8
C. 0, 1, 2, 3, 4, 5, 6, 7, 8
D. 1, 2, 3, 4, 5, 6, 7, 8
40. Se quiere realizar una encuesta en la que se quiere ver la cantidad de horas que personas de cierta ciudad dedican al deporte. La ciudad se caracteriza porque el 57% de su población son mujeres y el restante son hombres. Para la encuesta se requiere escoger una muestra representativa, ¿cuál de las siguientes muestras es la que mejor representa la población total de la ciudad?
- A. 57 mujeres y 86 hombres
B. 114 mujeres y 43 hombres
C. 150 mujeres y 150 hombres
D. 171 mujeres y 129 hombres
-

41. En un grupo de personas, 4 tienen 17 años, 5 tienen 18 años, 8 tienen 20 años, 4 tienen 22 años y 3 tienen 24 años. En este grupo se pueden hacer 231 parejas distintas. Si se hacen cada una de ellas y se calcula el promedio de sus edades. ¿Cuál será el valor promedio que más se repite?
- A. 18,5 años.
 - B. 19 años.
 - C. 21 años.
 - D. 22 años.

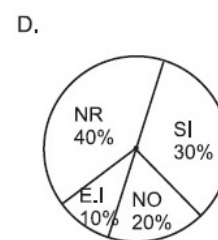
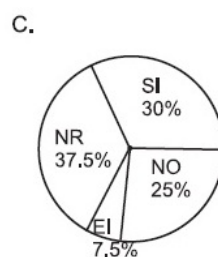
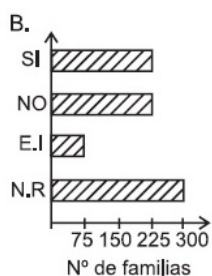
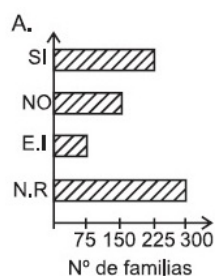
**RESPONDA LAS PREGUNTAS 42 Y 43 DE ACUERDO CON LA
SIGUIENTE INFORMACIÓN**

Para tomar la decisión de construir una plaza de mercado en el barrio Los Rosales, la Junta de Acción Comunal desea contar con el apoyo de la mayoría de las familias que allí viven. Para determinar qué quiere la mayoría, realizaron un sondeo en el que preguntaron: “¿Cree usted que sería de beneficio para el sector la construcción de una plaza de mercado?”. Los resultados se muestran en la siguiente tabla:

Respuesta	No. de familias
Si	225
No	150
Esta inseguro	75
No respondió	300

42. La Junta de Acción Comunal se inclinó por NO construir una plaza de mercado, debido a que los resultados del sondeo muestran que
-

- A. el 70 % de familias encuestadas no respondió afirmativamente.
- B. la mitad de familias encuestadas estuvieron inseguras o no respondieron la encuesta
- C. el número de familias que respondieron “sí”, supera a quienes respondieron negativamente en un 50 %
- D. el número de familias que respondieron “no” es el doble de las que están inseguras
43. Un gráfico que se podría presentar a los habitantes del barrio, sobre los resultados del sondeo, es

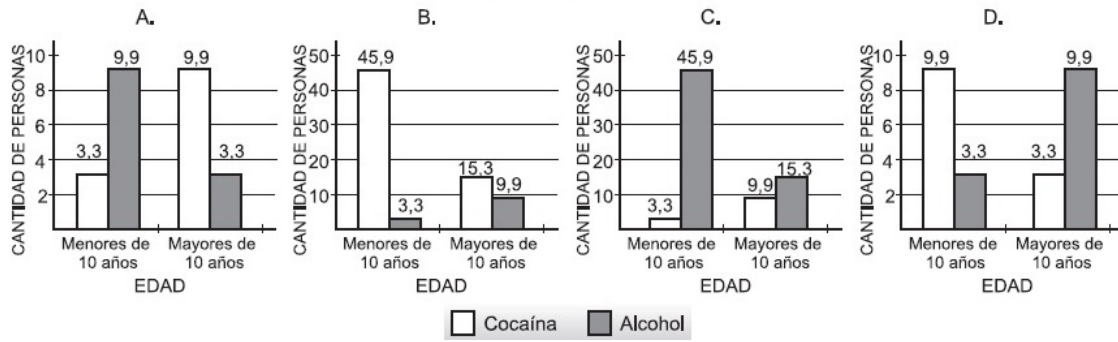


RESPONDA LAS PREGUNTAS 44 A 46 DE ACUERDO CON LA SIGUIENTE INFORMACIÓN

En el siguiente texto, se proporciona información sobre una investigación llevada a cabo, entorno a adicciones: “. . . en una muestra de 120 indigentes de corta edad [. . .] se constató que únicamente en el mes anterior a la consulta, 86 % de los muchachos habían consumido tabaco, 51 % alcohol, 44 % marihuana, 11 % cocaína y 56 % inhalantes. Además 26 de ellos afirmaron haber ingerido drogas farmacéuticas”.

44. Un antropólogo, que adelantó una investigación sobre el mismo tema, lee el texto y toma algunos apuntes útiles para su estudio; sin darse cuenta, hace una interpretación errada del texto, esta es:

-
- A. más del 30 % de los jóvenes examinados habían consumido tabaco y alcohol, un mes antes a la consulta.
- B. un mes antes a la consulta, los 120 jóvenes habían consumido inhalantes o marihuana.
- C. un mes antes a la consulta, el 7 % de los jóvenes consumieron inhalantes y alcohol.
- D. el consumo de cocaína, un mes antes a la consulta, fue menor al de otras sustancias, incluso al de drogas farmacéuticas.
45. Tomando como fuente el texto presentado, un periodista ha preparado un artículo en el que afirma que el 30 % de los muchachos consumió, un mes antes a la consulta, drogas farmacéuticas. Antes de ser publicado el artículo, se le sugiere que cambie esta afirmación, porque
- A. no fué la tercera parte de la muestra, la que consumió drogas farmacéuticas un mes antes a la consulta.
- B. estaría incluyendo a 10 personas que no consumieron drogas farmacéuticas un mes antes a la consulta.
- C. estaría incluyendo a 6 personas que no consumieron drogas farmacéuticas un mes antes a la consulta.
- D. no fueron 30 personas las que consumieron drogas farmacéuticas un mes antes a la consulta Tomando como fuente el texto presentado.
46. Profundizando en el estudio, se encontró que la cuarta parte de los jóvenes que consumieron cocaína, eran menores de 10 años mientras que la cuarta parte de los jóvenes que consumieron alcohol eran mayores de 10 años. Estos resultados pueden presentarse al público mediante el gráfico
-



Bibliografía

- [1] Quintero Pérez, Luis Eduardo, *Competencias y estándares en Matemáticas*, Los Tres editores S.A., Cali, 2009.
- [2] MARTÍNEZ BENCARDINO CIRO, *Estadística Básica Aplicada*, editorial Ecoe Ediciones, Bogotá, 2001.
- [3] URIBE JULIO A. *Matemática, una propuesta curricular 11*, editorial Bedout. Medellín, 1990.
- [4] PEÑA ANGELA JULIETA, *Inteligencia Lógico matemática 11*, editorial voluntad. Bogotá 2003.
- [5] DÍAZ D. RICARDO, *Pensamiento Matemático 10*, editorial Libros & Libros. Bogotá 2003.
- [6] VIDALES LUIS, *Historia de la estadística en Colombia*. Bogotá 1978.