Colegio de Estudios Científicos y Tecnológicos del Estado de Guanajuato.

PROBABILIDAD Y ESTADÍSTICA

CUADERNO DE TRABAJO
SEXTO SEMESTRE

PROBABILIDAD ESTADÍSTICA

TEMAS DE FILOSOFÍA

ASIGNATURA PROPEDÉUTICA

Número de registro:
03-2021-121412514400-01

Secretaría de Educación de Guanajuato

EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA

Mensaje de la Directora General

Joven Estudiante:

En todo este proceso de incorporación al mundo profesional, las matemáticas tienen una importancia decisiva, por lo que su aprendizaje en la preparatoria es de la mayor importancia. Veamos por qué.

La competencia lógico matemática, la capacidad de escuchar; la expresión oral clara y la redacción lógica nos permiten incorporar información nueva y transmitirla en cualquier situación, sea escolar o laboral. Estas habilidades son, por lo tanto, la puerta de entrada para conocer todo lo que nos rodea (incluso las demás disciplinas) y para darnos a conocer a quienes nos rodean. Sin estas habilidades básicas no podemos tener éxito en la vida social adulta.

La reflexión sobre el uso cotidiano y su mejor conocimiento conducen aun pensamiento más ordenado, por lo que el aprendizaje de las materias básicas en la preparatoria permite a los alumnos tener un instrumento para clasificar mejor sus ideas.

En todo acto de comunicación, ya sea símbolos, números, de forma oral o escrita, intervienen una serie de elementos necesarios para que dicho acto sea eficaz. O lo quees lo mismo, sin estos componentes el proceso comunicativo no sería posible.

Directorio

Dra. Virginia Aguilera Santoyo
Directora General

Ing. Miguel Espartaco Hernández García
Encargado de la Dirección Académica

C.P Vicenta Martínez Torres
Directora Financiera y Administrativa

Lic. Sara Cecilia Casillas Martínez
Directora de Planeación y Desarrollo

Lic. Carlos Alberto Gorostieta Romero
Director de Vinculación

C.P. Alfredo García Flores
Director de Desarrollo Humano

Lic. Jaime Díaz Zavala
Director de Asuntos Jurídicos

LIA. Reynaldo Nava Garnica
Subdirector de Sistemas e informática Educativa

Comité Editorial

Docentes Participantes

Cuaderno de Trabajo de Probabilidad y Estadística

Gerardo Medina Jiménez - Plantel Comonfort.

Néstor José Guevara Ordoñez - Plantel León.

José de Jesús Leos Mireles - Plantel Silao.

Olga Lucía Guerrero Zepeda - Plantel Celaya II.

Mariana Hernández Hernandéz - Plantel San Juan de la Vega.

Antonio Ciénega - Plantel León I.

CONTENIDO

Aprendiendo a usar el cuaderno:

Símbolos de Identificación

 **Rescatando mis Aprendizaje.**

 **Para aprender más**

 **Ejercitando mi habilidad.**

 **¿Qué Aprendí?**

 **Rescatando mis Aprendizaje**

 **Actividad Transversal**

INTRODUCCIÓN.

Propósito del tema.

Que el estudiante:

1.-Analice fenómenos sociales o naturales, utilizando las herramientas básicas de la estadística descriptiva, para muestra, procesar y comunicar información social y científica, para la toma de decisiones.

2.-Exprese Razones para fundamentar una respuesta y obtenga conclusiones pertinentes a partir de datos estadísticos.

3.-Identifique e interprete los datos a partir de una representación gráfica de datos.

Competencias disciplinares

1. Construye e interpreta modelos matemáticos, mediante la aplicación de procedimientos aritméticos, algebraicos, geométricos y variacionales, para la comprensión y análisis de situaciones reales, hipotéticas o formales.
3. Explica e interpreta los resultados obtenidos mediante procedimientos matemáticos y los contrasta con modelos establecidos o situaciones reales.
5. Analiza las relaciones entre dos o más variables de un proceso social o natural para determinar o estimar su comportamiento.
8. Interpreta tablas, graficas, mapas, diagramas y textos con símbolos matemáticos y científicos.

Estructura del Cuadro de Contenidos de Probabilidad y Estadística

Eje	Componente	Contenido central	Contenidos específicos	Aprendizajes esperados	Producto esperado
Del manejo de la información al pensamiento estocástico	Riesgo, inferencia y aleatoriedad: Elementos de la Estadística y la Probabilidad.	Conceptos básicos d Estadística y Probabilidad. Recolección de datos y su clasificación en clases. Uso del conteo y la probabilidad para eventos.	• Nociones y conceptos básicos de estadística y probabilidad. • Enfoques de probabilidad. ¿Qué significa cada enfoque de probabilidad?, ¿qué significan las medidas de tendencia central?, ¿para qué obtener estos valores? Técnicas de conteo y agrupación en clases para la determinación de probabilidades.	• Usa un lenguaje propio para situaciones que necesiten del estudio con elementos de estadística y probabilidad. • Usa técnicas de conteo o agrupación en la determinación de probabilidades. • Organiza la información como parte de la estadística para el estudio de la probabilidad. • Estudia el complemento que ofrece la estadística para la probabilidad.	• Cálculo del promedio de una colección de datos.
Del manejo de la información al pensamiento estocástico	Riesgo, inferencia y aleatoriedad: Elementos de la Estadística y la Probabilidad	Concepto de riesgo en situaciones contextuales. Contextualización de los elementos de probabilidad condicional e interpretación intuitiva del teorema de Bayes (probabilidad subjetiva).	• ¿Qué es el riego?, ¿qué papel juega la probabilidad y estadística en el estudio del riesgo? • Usos de la estadística y probabilidad en situaciones dadas. • Análisis de la información. • Nociones de incertidumbre, azar y aleatoriedad. • Tipos de eventos en el estudio de la probabilidad.	• Reconoce la diversidad de situaciones que precisan de la incertidumbre en el tratamiento del riesgo. • Modela con estadística y probabilidad el estudio de la información. • Organiza la información recolectada de la situación estudiada. Construye fórmulas de probabilidad.	• Construcción de tablas de frecuencia. • Cálculo de la probabilidad de un evento.

Eje	Componente	Contenido central	Contenidos específicos	Aprendizajes esperados	Producto esperado
Del manejo de la información al pensamiento estocástico.	Riesgo, inferencia y aleatoriedad: Elementos de la Estadística y la Probabilidad.	Manejo de la Información en situaciones de la vida cotidiana.	• Estudio de la información. ¿Qué papel juegan las medidas de tendencia central?, ¿cómo representar la información en un gráfico estadístico?, ¿cómo estudiar un gráfico estadístico?, ¿qué papel juega la probabilidad en el manejo de la información? • Cálculo de las medidas de tendencia central y su representatividad en términos de la variabilidad y contexto situacional. • Construcción de gráficos estadísticos en la representación de la información. Análisis de tipos de gráficos estadísticos.	• Recolecta y ordena la información de alguna situación. • Interpreta y analiza la información. • Representa la información. • Toma decisiones a partir del análisis de la información.	• Construcción de distintos tipos de gráficos y emisión de opiniones derivadas de ellos.
Del manejo de la información al pensamiento estocástico.	Riesgo, inferencia y aleatoriedad: Elementos de la Estadística y la Probabilidad.	Tratamiento de las medidas de tendencia central. Tratamiento y Significado de medidas de Dispersión.	• Medidas de tendencia central. ¿Qué es la moda, la media aritmética, la mediana? ¿Qué es un cuartil?, ¿qué es una medida de dispersión?, ¿qué es una medida de forma?, ¿qué es una medida de correlación? • Análisis de la información y toma de decisiones. ¿Qué información brindan las • medidas de tendencia central?, ¿Cuándo se puede considerar que todas dan la misma información? ¿en cualquier fenómeno tiene significado?	• Calcula las medidas de tendencia central, medidas de dispersión, medidas de forma y medidas de correlación. • Interpreta las medidas de tendencia central desde el análisis del gráfico estadístico, así como su variabilidad y representación de la situación contextual. Toma decisiones a partir de las medidas de tendencia central y su representación con respecto a un conjunto de datos.	• Argumento de qué es una medida de tendencia central y qué es una medida de dispersión. • Ejemplos de dichas medidas. • Construcción de cuartiles a partir de datos dados.

UNIDAD I

UNIDAD I.

El Cuaderno de Probabilidad y Estadística refleja de forma sencilla y práctica los contenidos conforme al Plan de Estudios vigente, así mismo han quedado plasmadas las inquietudes de los docentes del Comité Técnico de la Academia Estatal de Matemáticas, quienes pusieron el mejor empeño para que tú, estudiante, cuentes con herramientas fundamentales para el termino exitoso de esta asignatura.

En esta unidad aprenderás entre otros temas las Técnicas de conteo, Conjuntos, Diagramas de conjuntos Conceptos de estadística y fundamentos de Probabilidad y verás también la Teoría de Bayes.

Al término del curso el alumno será capaz de:

- Aplicar el lenguaje correcto y específico de la asignatura.
- Comprende la necesidad y oportunidad de aplicar modelos estadísticos no sólo en la ciencia sino también en la tecnología y en las distintas ramas del saber.
- Comprende las posibilidades, ventajas y limitaciones de los modelos estadísticos, como simple modelo de una realidad, basado en la matemática, o una ciencia formal, y no como la realidad misma.
- Aplica y evalúa los principios estadísticos para resolver problemas generales.

Rescatando mis Aprendizajes.

DIAGNÓSTICO TÉCNICAS DE CONTEO.

Instrucciones: Resuelve los siguientes planteamientos en base a tu experiencia y conocimiento en los temas, recuerda traer a flor de piel lo que sabes.

1) Alan pretende dar aventón en su motocicleta a sus amigos Bertha, Carlos y Diana, si se sabe que solamente pueden ir 3 a la vez en la motocicleta, contesta;
   a) ¿De cuantas maneras pueden ir en la motocicleta, sin restricción alguna?
   b) ¿De cuantas maneras pueden repartirse en la motocicleta, si se toma en cuenta el orden?
2) ¿De cuantas maneras pueden repartirse tres premios a diez personas, si cada persona solo puede obtener un premio?
3) ¿De cuantas maneras puede elegir su carrera un aspirante a estudiante en CECyTE Guanajuato, de entre las especialidades de Administración, Informática, Mantenimiento Industrial, Instrumentación y Derecho?
4) Describe el conjunto de los números naturales menores a diez:
5) Si tenemos los conjuntos; A = {2, 3, 5, 7, 11, 13, 17, 19} y B = {5, 10, 15, 20}, contestar:
   a) ¿Cuál es el resultado si unimos ambos conjuntos?
   b) ¿Existen elementos que se encuentren en ambos conjuntos?
   c) ¿Cuáles elementos son?
6) ¿Cuál es el número de resultados posibles que se pueden dar al lanzar un par de monedas corrientes?

Para aprender más

DEFINICIONES Y APLICACIONES DE LA ESTADÍSTICA

Para iniciar, se presentan algunas definiciones complementarias de la estadística:

a) Es el la ciencia o arte de reunir y analizar e inferir consecuencias a partir de estos elementos.

b) Es la rama de las matemáticas que aborda los datos numéricos o cuantitativos y los relaciona con el método científico en la toma, organización, recopilación, presentación y análisis de los mismos, con el fin de tomar decisiones responsables.

c) Es la ciencia que se ocupa de la reunión o recopilación de todos los hechos que se pueden valorar mecánicamente para hacer comparaciones entre las cifras y obtener conclusiones, aplicando la teoría de las probabilidades.

d) Es un conjunto de métodos o técnicas que estudia y analiza los datos que son susceptibles de expresión numérica para llegar a conclusiones que permitan tomar decisiones y pronosticar las consecuencias de las mismas.

c) El estudio de los fenómenos al azar, que agrupa, clasifica y ordena experiencias y observaciones sobre la manifestación de hechos, para deducir las leyes o los principios que los rigen.

La Estadística descriptiva es la parte de la estadística que también se conoce con el nombre de estadística deductiva, ya que trata solamente de describir y analizar un grupo de datos dados, los cuales se representan con tablas, graficas, cuadros e índices. No realiza inferencias y, por tanto, no obtiene conclusiones a partir de un grupo mayor de datos.

Ejemplos

Cuando un investigador conduce un estudio, reúne gran cantidad de información numérica o datos acerca del problema en cuestión. Los datos pueden tener formas variadas, por ejemplo: datos de frecuencia (recuentos de votantes que prefieren uno u otro candidato político) o datos escolares (los cocientes intelectuales de un grupo de estudiantes, los pesos netos de los ingredientes para hacer un pastel). Para realizar la función descriptiva, el estadista formula reglas y procedimientos para la presentación de los datos en forma más útil y significativa, como los gráficos y los parámetros. Otro ejemplo es cuando el jefe de personal de una empresa aplica una prueba de habilidades y destrezas a un grupo de trabajadores; utilizando estadística descriptiva en las puntuaciones obtenidas, los datos se tienen que ordenar, clasificar, calcular promedios, identificar las cualidades más típicas, construir tablas, gráficas y cuadros para comparaciones, predecir el rendimiento.

Elementos de la estadística

Entre los elementos básicos de la estadística, se encuentran fundamentalmente la población o el universo, la variable (continua y discreta) y los tipos de datos, todos ellos con el fin de fortalecer la comprensión de los procedimientos estadísticos.

Recopilación de datos

Los datos son situaciones o hechos que representan numéricamente y que algunas veces forman parte de la vida cotidiana y otras, se encuentran en libros por que han sido recopilados anteriormente por otras personas.

Manejo de los datos

Una vez hecha la recopilación de datos, el siguiente paso es su correcta organización para que brinden una información fiel y de gran utilidad, considerando su naturaleza de acuerdo con los propósitos para los que fueron recopilados.

Redondeo de datos

El redondeo de datos es una técnica útil, ya que tiene la finalidad de reducir el mínimo de errores acumulados por dicha práctica, sobre todo cuando la cantidad de datos es muy grande.

Ejemplo: Suma los datos 7.75, 9.85. 4.55, 13.65, 8.45, 10.35, 12.15 en las siguientes formas

a) Directa

b) Redondeando por el criterio del número par más próximo.

c) Redondeando los datos por exceso.

Solución.

a)	b)	c)
7.75	7.8	7.8
9.85	9.8	9.9
4.55	4.6	4.6
13.65	13.6	13.7
8.45	8.4	8.5
10.35	10.4	10.4
12.15	12.2	12.2
66.65	66.8	67.1

Se hace notar que el resultado en b) es más exacto que el que se obtiene en c), ya que se acerca más el resultado directo en a); también porque la acumulación de error por redondeo es mucho menor al aplicar el criterio del número par más próximo que antecede o es consecuente del 5.

Rescatando mis Aprendizaje.

Evaluación diagnóstica

Instrucciones. - Realiza lo que se te indica a continuación.

1. Indica que variables son cualitativas y cuales cuantitativas.

   a) Comida favorita.
   b) Profesión que te gusta.
   c) Numero de penales anotados por ti en la última temporada.
   d) Numero de sillas en tu escuela.
   e) El color de los autos del estacionamiento de la escuela

2. Clasificar las siguientes variables en cualitativas y cuantitativas discretas y continuas.

   a) La nacionalidad de una persona.
   b) Numero de litros de agua de la reserva de tu casa.
   c) Numero de medicamentos en un estanque de farmacia.
   d) La profesión de una persona.
   e) El área de un edificio.

3. A un conjunto de 5 números cuya media es 7.31 se le añaden los números 4.47 y 10.15. ¿Cuál es la media del nuevo conjunto de números?

Para aprender más

ESTADÍSTICA DESCRIPTIVA.

La estadística descriptiva comprende las técnicas que se emplean para resumir y describir datos numéricos. Son sencillas desde el punto de vista matemático y su análisis se limita a los datos coleccionados sin inferir en un grupo mayor.

El estudio de los datos se realiza con representaciones gráficas, tablas, medidas de tendencia central y dispersión.

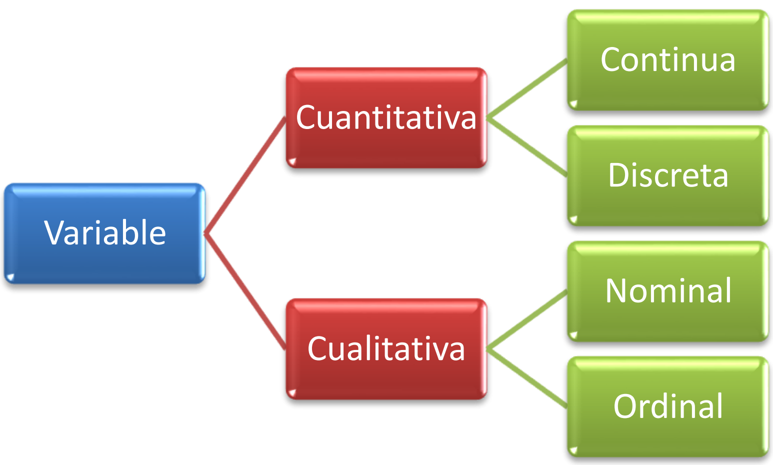
Estadística inferencial. El problema crucial de la estadística inferencial es llegar a proposiciones acerca de la población a partir de la observación efectuada en muestras bajo condiciones de incertidumbre. Ésta comprende las técnicas que, aplicadas en una muestra sometida a observación, permiten la toman de decisiones sobre una población o proceso estadístico. En otras palabras, es el proceso de hacer predicciones acerca de un todo basado en la información de una muestra. La inferencia se preocupa de la precisión de los estadígrafos descriptivos ya que estos se vinculan inductivamente con el valor poblacional.

- Población. Es el conjunto de todos los elementos que presentan una característica común determinada, observable y medible. Por ejemplo, si el elemento es una persona, se pueden estudiar las características de edad, peso, nacionalidad, sexo, etc. Los elementos que integran una población pueden corresponder a personas, objetos o grupos (por ejemplo, familias, fábricas, empresas, entre otros). Las características de la población se resumen en valores llamados parámetros.

- Muestra. La mayoría de los estudios estadísticos, se realizan no sobre la población, sino sobre un subconjunto o una parte de ella llamada muestra, partiendo del supuesto de que este subconjunto presenta el mismo comportamiento y características que la población. En general el tamaño de la muestra es mucho menor al tamaño de la población. Los valores o índices que se concluyen de una muestra se llaman estadígrafos y estos mediante métodos inferenciales o probabilísticos, se aproximan a los parámetros poblacionales.

- Variable. Se llama variable a una característica que se observa en una población o muestra, y a la cual se desea estudiar. La variable puede tomar diferentes valores dependiendo de cada individuo. Una variable se puede clasificar de la siguiente manera:

Imagen No. 1 Distribución de las variables



Fuente: Elaboración propia, noviembre 2020

- Variable cuantitativa. Es aquella que toma valores numéricos. Dentro de ella, se subdividen en:
  - Continua. Son valores reales. Pueden tomar cualquier valor dentro de un intervalo, por ejemplo, el Peso, estatura, sueldos, metros de un pintarrón, entre otros.
  - Discreta. Toma valores enteros. Por ejemplo, el número de hijos de una familia, número de alumnos de un curso, número de libros en un estante, número de bicicletas en un estacionamiento, entre otros.

- Variable cualitativa. Es aquella que describe cualidades. No son numéricas y se subdividen en:
  - Nominal. Son cualidades sin orden. Por ejemplo, el estado civil, preferencia por una marca, sexo, lugar de residencia.
  - Ordinal. Son cualidades que representan un orden y jerarquía. Por ejemplo, el nivel educacional, días de la semana, calidad de la atención, nivel socioeconómico, escalas evaluativas en una prueba.

- Obtención de los datos. La obtención de la información se puede realizar por diversos medios. Una forma es a través de una encuesta a un grupo de individuos, donde a cada uno se le hacen las mismas preguntas. Otra forma es a través de experimentos donde la respuesta a la variable es el resultado del experimento. Puede también recolectarse los datos en forma directa, es decir, la información se extrae de alguna base de datos seleccionando una muestra de ellos.

En cualquiera de estos casos contamos con una selección de información llamada **muestra** y que se procede a analizar. Existen diferentes técnicas para realizar el muestreo y que dependerán cada caso, cual usar. Algunas de ellas son:

- ✓ Muestreos aleatorios simple. Todos los elementos de la población tienen igual posibilidad de ser escogidos y se eligen al azar.

- ✓ Muestreo sistemático. Los elementos se seleccionan a un intervalo uniforme en una lista ordenada. Una preocupación del muestreo sistemático es la existencia de factores cíclicos en el listado que pudieran dar lugar a un error.

- ✓ Muestreo estratificado. Los elementos de la población son primeramente clasificados en grupos o estratos según una característica importante. Luego, de cada estrato se extrae una muestra aleatoria simple.

- ✓ Muestreo por conglomerado. Los elementos de la población están subdivididos en grupos y se extraen aleatoriamente algunos de estos grupos completos.

Ejercitando mi habilidad.

Instrucciones. - Realiza los siguientes ejercicios de clasificación de variables en cualitativas o cuantitativas y sus tipos continuas, discretas, nominal u ordinal.

1.- El tipo de sangre de los habitantes de la Ciudad de León Guanajuato.

2.- Los posibles valores de variable son: A+, A−, O+, O−, AB+, AB−, B+ y B−.

3.- El deporte que practican con mayor frecuencia los estudiantes de quinto semestre de la especialidad de Mantenimiento Industrial.

4.- Los posibles valores de variable son: Fútbol, Basquetbol, Volibol, Natación, Béisbol y atletismo.

5.- El medio de comunicación al que recurren los padres de familia de los jóvenes del Plantel Irapuato 1 para enterarse de las noticias.

6.- Los posibles valores de variable son: Televisión, radio, prensa escrita, internet, entre otros.

7.- El grado escolar de los empleados de una empresa de una comercializadora.

8.- Los posibles valores son: Primaria, Secundaria, Bachillerato, Licenciatura, Maestría y Doctorado.

9.- El día de la semana en que aplicaron las encuestas para el INEGI.

10.- Los posibles valores de variable: Domingo, lunes, martes, miércoles, jueves, viernes y sábado.

11.- El Semestre que cursa un estudiante universitario.

12.- Los posibles valores de variable: Primero, segundo, tercero, cuarto, etc.

13.- El número de hermanos que tiene cada estudiante del grupo.

14.- La cantidad de alumnos que egresan del Plantel Celaya cada año.

15.- La talla de calzado de cada uno de los integrantes de una familia.

16.- La suma de puntos al lanzar dos dados.

17.- El número de águilas que caen al lanzar tres monedas corrientes.

18.- La estatura de los alumnos de un grupo (depende de la precisión del instrumento utilizado para medir longitudes).

19.- La temperatura ambiental cada hora durante todo un día, en uno de los 72 Municipios de nuestro Estado de Guanajuato (dependerá de la precisión del termómetro que se utilice).

20.- El nivel del agua que registra la presa la cuenca de la Esperanza.

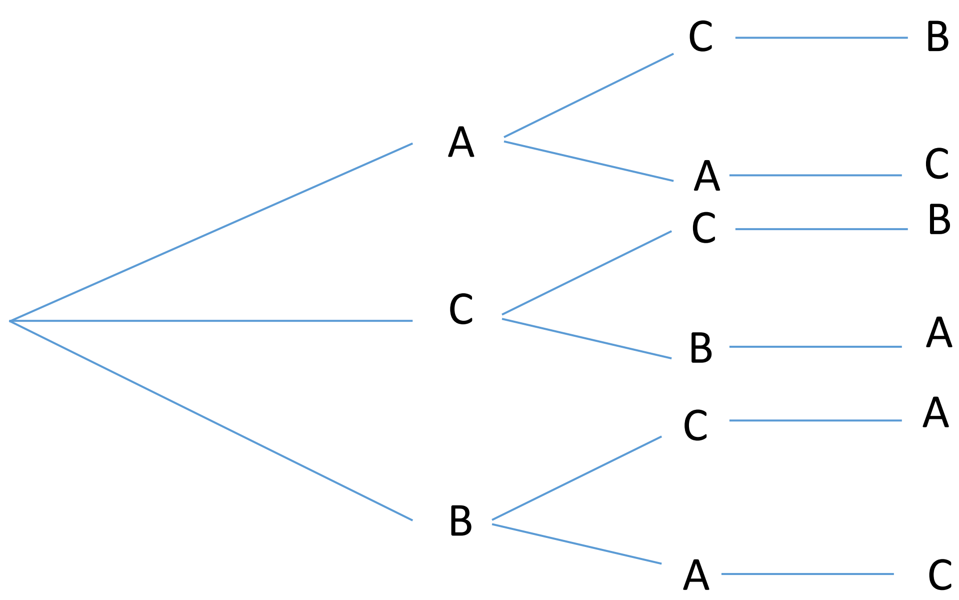
Para aprender más

DIAGRAMA DE ÁRBOL.

Los diagramas de árbol son mecanismos empleados para enumerar todas las posibilidades de una secuencia de eventos, donde cada evento puede ocurrir en un número finito de formas. Los diagramas de árbol proporcionan un método sistemático de enumeración objetiva de los resultados.

Como cuando se tienen tres libros, uno de álgebra, otro de contabilidad y otro de biología ¿De cuantas formas distintas se puede ordenar los libros?

Imagen No. 2 Diagrama de árbol



Fuente: Elaboración propia, noviembre 2020

La solución es {ACB, ABC, BCA, CAB, CBA}, esto es que se tienen 6 formas distintas para acomodar los tres libros.

Para construir un diagrama de árbol se inicia con una rama para cada una de las posibilidades. Al final de cada rama se constituye un **nodo** del cual parten nuevas ramas, según las posibilidades del siguiente paso, tantas veces como se requiera hasta llegar al final del experimento.

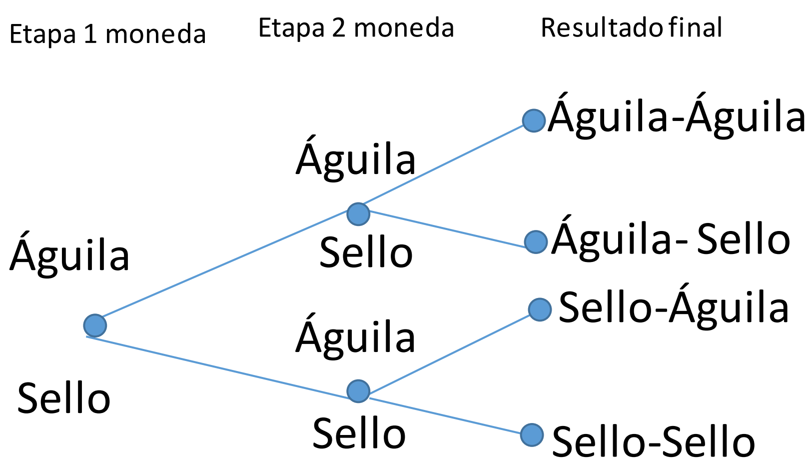
En la siguiente figura se muestra un experimento de lanzar dos monedas, la secuencia de pasos es de izquierda a derecha.

Paso 1: Corresponde a lanzar la primera moneda y tiene 2 ramificaciones que corresponden a 2 resultados (águila o sello)
Paso 2: Corresponde a lanzar la segunda moneda y para cada resultado posible en el paso 1 tiene 2 ramas que corresponden a los 2 resultados posibles del paso 2

Por último, cada uno de los puntos en el extremo derecho del árbol corresponden a un resultado experimental.

Cada trayectoria a través del árbol desde el primer nodo a uno de los nodos de la derecha del árbol corresponde a una secuencia única de resultados.

Imagen No. 3 Ejemplo de Diagrama de árbol



Fuente: Elaboración propia, noviembre 2020

Conclusión

Como podrás observar, se tienen 4 resultado s finales (águila-águila) (águila-sello) (sello-águila) y (sello-sello).

CONJUNTOS.

Se entiende por conjunto una colección de una clase particular o cualquier agrupación o colección de objetos bien definidos, de tal manera que se puede decir si un objeto pertenece o no al conjunto. Por lo general, se presenta con letras mayúsculas (A, B, C, D, entre otras).

Un elemento de un conjunto es cada uno de los miembros que lo constituyen. Los elementos de un conjunto se representan mediante alguna notación particular que se acuerde para nombrarlos, puede ser con letras minúsculas (a, b, c, d, etcétera).

OPERACIÓN DE UNIÓN ENTRE CONJUNTOS.

Si $A \cup B = \emptyset$, esto es, si A y B no tienen elementos en común, entonces se dice que A y B son mutuamente excluyentes, disjuntos o ajenos.

Sean A y B conjuntos arbitrarios.

La Unión de A y B, expresada por $A \cup B$, es el conjunto de elementos que pertenecen a A y/o B:

Ejemplo 1: $A = \{1,2,3\}$ y $B = \{2,4,6\}$, entonces $A \cup B = \{1,2,3,4,5,6\}$

Ejemplo 2: $A = \{1,2,3\}$ y $B = \{2,4,5\}$, entonces $A \cup B = \{1,2,3,4,5\}$

OPERACIÓN DE INTERSECCIÓN ENTRE CONJUNTOS.

La Intersección de A y B, expresada por $A \cap B$, es el conjunto de elementos comunes a A y a B:

Ejemplo 1: A y B son ajenos, $A = \{1,2,3\}$, $B = \{4,5,6\}$, entonces $A \cap B = \emptyset$

Ejemplo 2: A y B no son ajenos, $A = \{1,2,3\}$, $B = \{2,4,5\}$, entonces $A \cap B = \{2\}$.

PRINCIPIO DE MULTIPLICACIÓN.

Si un evento A puede ocurrir de n(A) maneras diferentes, y una vez que este ha ocurrido, otro evento B puede ocurrir de n(B) maneras diferentes, entonces el número total de formas diferentes en que ambos eventos pueden ocurrir en el orden indicado (AXB), es igual a n(A) X n(B).

Ejemplo 1; ¿De cuantas maneras pueden repartirse 5 premios a 20 personas, si cada persona solo puede obtener un premio?

Aplicando el principio fundamental del conteo, tenemos 20 personas que pueden recibir el primer premio. Una vez que éste ha sido entregado, restan 19 personas para recibir el segundo, posteriormente quedarán 18 personas para el tercer premio, de ahí 17 personas para el cuarto y finalmente 16 para el quinto premio.

De ahí que el número de maneras distintas de repartir los cinco premios es 20X19X18X17X16=1,869,480 formas.

PRINCIPIO DE SUMA.

Si un evento A puede ocurrir de n(A) maneras diferentes, y una vez que este ha ocurrido, otro evento ajeno B puede ocurrir de n(B) maneras diferentes, entonces el número total de formas diferentes en que los eventos A o B pueden ocurrir en el orden indicado (A∩B) = n(A) + n(B) maneras distintas.

Ejemplo 1; Una pareja de recién casados que han reunido el dinero suficiente para dar el enganche de una casa, tiene las siguientes alternativas:

F1 = Fraccionamiento Jardines del Real: Tipo colonial, tipo provenzal francés y tipo californiano.

F2 = Fraccionamiento Villa del Viento: Tipo residencial y tipo económico.

¿Cuántas opciones tiene la pareja para elegir?

n(F1∩F2) = n(F1) + n(F2) = 3 + 2 = 5 opciones.

PERMUTACIÓN.

Podemos decir que las permutaciones de un conjunto de elementos, es un ordenamiento específico de todos o algunos de esos elementos del conjunto, las permutaciones facilitan el recuento de las diferentes disposiciones ordenadas que pueden hacerse con los elementos del conjunto. En una permutación es muy importante considerar el orden en que se disponen los elementos del conjunto.

Recurriendo al principio de la multiplicación o principio fundamental del conteo podemos enunciar que el número de permutaciones P de n objetos distintos tomados de r en r, es:

Ejemplo 1; Imaginemos que 5 personas desean nombrar un comité directivo compuesto de un presidente, un vicepresidente, un secretario y un tesorero ¿Cuántas maneras distintas existen para conformar dicho comité?

P (5, 4) = 5! / (5-4)! = 5! / 1! = 1X2X3X4X5 / 1 = 120 maneras.

COMBINACIÓN.

En una permutación el orden de los elementos es vital, pero cuando el orden en que se disponen los elementos carece de importancia, a esta disposición de elementos le denominamos combinación.

Por lo tanto, una combinación es un subconjunto o una disposición de todos los elementos de un conjunto, sin tener en cuenta el orden de ellos. El número de combinaciones o subconjuntos no ordenados, en donde cada uno está formado por r elementos, que se obtienen de un conjunto de n elementos, está dada por:

Ejemplo 1; Supongamos que queremos tomar 2 de 3 libros, ¿Cuántas combinaciones pueden realizarse entre estos 2 libros?

C (3, 2) = 3! / (3-2)! 2! = 3! / 1! 2! = 1X2X3 / 1X2 = 6 / 2 = 3 combinaciones.

COMBINACIONES Y PERMUTACIONES.

¿Qué diferencia hay?

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 pusimos 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**.

Así que en matemáticas usamos un lenguaje más preciso:

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

	¡Así que lo de arriba se podría llamar "cerradura de permutación"!

Con otras palabras:

Una permutación es una combinación **ordenada**.

Para ayudarte a recordar, piensa en "**P**ermutación... **P**osición"

PERMUTACIONES

Hay dos tipos de permutaciones:

1. **Se permite repetir**: como la cerradura de arriba, podría ser "333".
2. **Sin repetición**: por ejemplo, los tres primeros en una carrera. No puedes quedar primero *y* segundo a la vez.

1. Permutaciones con repetición

Son las más fáciles de calcular. Si tienes ***n*** cosas para elegir y eliges ***r*** de ellas, las permutaciones posibles son:

$$n \times n \times \ldots \text{ (r veces)} = n^r$$

(Porque hay **n** posibilidades para la primera elección, DESPUÉS hay **n** posibilidades para la segunda elección, y así.)

Por ejemplo, en la cerradura de arriba, hay 10 números para elegir (0, 1...,9) y eliges 3 de ellos:

$$10 \times 10 \times \ldots \text{ (3 veces)} = 10^3 = 1000 \text{ permutaciones}$$

Así que la fórmula es simplemente:

$$n^r$$

donde ***n*** es el número de cosas que puedes elegir, y eliges ***r*** de ellas (Se puede repetir, el orden importa)

2. Permutaciones sin repetición

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

Por **ejemplo**, ¿cómo podrías ordenar 16 bolas de billar?

Después de elegir por ejemplo la "14" no puedes elegirla otra vez.

Así que tu primera elección tiene 16 posibilidades, y tu siguiente elección tiene 15 posibilidades, después 14, 13, etc. Y el total de permutaciones sería:

$$\mathbf{16 \times 15 \times 14 \times 13 \ldots = 20{,}922{,}789{,}888{,}000}$$

Pero a lo mejor no quieres elegirlas todas, sólo 3 de ellas, así que sería solamente:

$$\mathbf{16 \times 15 \times 14 = 3360}$$

Es decir, hay 3,360 maneras diferentes de elegir 3 bolas de billar de entre 16. ¿Pero cómo lo escribimos matemáticamente? Respuesta: usamos la "función factorial"

La **función factorial** (símbolo:!) significa que se multiplican números descendentes. Ejemplos:

- $4! = 4 \times 3 \times 2 \times 1 = 24$
- $7! = 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 5040$
- $1! = 1$

Nota: en general se está de acuerdo en que 0! **= 1**. Puede que parezca curioso que no multiplicar ningún número dé 1, pero ayuda a simplificar muchas ecuaciones.

Así que si quieres elegir **todas** las bolas de billar las permutaciones serían:

$$\mathbf{16! = 20{,}922{,}789{,}888{,}000}$$

Pero si sólo quieres elegir 3, tienes que dejar de multiplicar después de 14. ¿Cómo lo escribimos? Hay un buen truco, ¡dividimos entre 13!...

$$\frac{16 \times 15 \times 14 \times 13 \times 12 \ldots}{13 \times 12 \ldots} = 16 \times 15 \times 14 = 3360$$

¿Lo ves? $\mathbf{16! / 13! = 16 \times 15 \times 14}$

La fórmula se escribe:

$$\frac{n!}{(n-r)!}$$

donde ***n*** es el número de cosas que puedes elegir, y eliges ***r*** de ellas (No se puede repetir, el orden importa)

Ejemplos:

Nuestro "ejemplo de elegir en orden 3 bolas de 16" sería:

$$\frac{16!}{(16-3)!} = \frac{16!}{13!} = \frac{20{,}922{,}789{,}888{,}000}{6{,}227{,}020{,}800} = 3360$$

¿De cuántas maneras se pueden dar primer y segundo premio entre 10 personas?

$$\frac{10!}{(10-2)!} = \frac{10!}{8!} = \frac{3{,}628{,}800}{40{,}320} = 90$$

(que es lo mismo que: **10 × 9 = 90**)

Notación

En lugar de escribir toda la fórmula, la gente usa otras notaciones como:

$$P(n,r) = {}^{n}P_{r} = {}_{n}P_{r} = \frac{n!}{(n-r)!}$$

COMBINACIONES

También hay dos tipos de combinaciones (recuerda que ahora el orden **no** importa):

1. **Se puede repetir**: como monedas en tu bolsillo (5,5,5,10,10)
2. **Sin repetición**: como números de lotería (2,14,15,27,30,33)

1. Combinaciones con repetición

En realidad, son las más difíciles de explicar, así que las dejamos para luego.

2. Combinaciones sin repetición

Así funciona la lotería. Los números se eligen de uno en uno, y si tienes los números de la suerte (da igual el orden) ¡entonces has ganado!

La manera más fácil de explicarlo es:

- Imaginemos que el orden sí importa (permutaciones),
- Después lo cambiamos para que el orden **no** importe.

Volviendo a las bolas de billar, digamos que queremos saber qué 3 bolas se eligieron, no el orden.

Ya sabemos que 3 de 16 dan 3360 permutaciones.

Pero muchas de ellas son iguales para nosotros, porque no nos importa el orden.

Por ejemplo, digamos que se tomaron las bolas 1, 2 y 3. Las posibilidades son:

El orden importa	El orden no importa
1 2 3 1 3 2 2 1 3 2 3 1 3 1 2 3 2 1	1 2 3

Así que las permutaciones son 6 veces más posibilidades.

De hecho, hay una manera fácil de saber de cuántas maneras "1 2 3" se pueden ordenar, y ya la sabemos. La respuesta es:

3! = 3 × 2 × 1 = 6

(Otro ejemplo: 4 cosas se pueden ordenar de **4! = 4 × 3 × 2 × 1 = 24** maneras distintas, ¡prueba tú mismo!)

Así que sólo tenemos que ajustar nuestra fórmula de permutaciones para **reducir** por las maneras de ordenar los objetos elegidos (porque no nos interesa ordenarlos):

$$\frac{n!}{(n-r)!} \times \frac{1}{r!} = \frac{n!}{r!(n-r)!}$$

Esta fórmula es tan importante que normalmente se la escribe con grandes paréntesis, así:

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

donde ***n*** es el número de cosas que puedes elegir, y eliges ***r*** de ellas
(No se puede repetir, el orden no importa)

Y se la llama "coeficiente binomial".

Notación

Además de los "grandes paréntesis", la gente también usa estas notaciones:

$$C(n,r) = {}^{n}C_{r} = {}_{n}C_{r} = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Ejemplo

Entonces, nuestro ejemplo de bolas de billar (ahora sin orden) es:

$$\frac{16!}{3!(16-3)!} = \frac{16!}{3! \times 13!} = \frac{20{,}922{,}789{,}888{,}000}{6 \times 6{,}227{,}020{,}800} = 560$$

O lo puedes hacer así:

$$\frac{16 \times 15 \times 14}{3 \times 2 \times 1} = \frac{3360}{6} = 560$$

Así que recuerda, haz las permutaciones, después reduce entre "**r!**" ... o mejor todavía... ¡Recuerda la fórmula!

Es interesante darse cuenta de que la fórmula es bonita y **simétrica**:

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

Con otras palabras, elegir 3 bolas de 16 da las mismas combinaciones que elegir 13 bolas de 16.

$$\frac{16!}{3!(16-3)!} = \frac{16!}{13!(16-13)!} = \frac{16!}{3! \times 13!} = 560$$

1. Combinaciones con repetición

Digamos que tenemos cinco sabores de helado: **banana, chocolate, limón, fresa y vainilla**. Puedes tomar 3 paladas. ¿Cuántas variaciones hay?

Vamos a usar letras para los sabores: {b, c, l, s, v}. Algunos ejemplos son

- {c, c, c} (3 de chocolate)
- {b, l, v} (uno de banana, uno de limón y uno de vainilla)
- {b, v, v} (uno de banana, dos de vainilla)

(Y para dejarlo claro: hay **n=5** cosas para elegir, y eliges **r=3** de ellas. El orden no importa, ¡y **sí** puedes repetir!)

Bien, no puedo decirte directamente cómo se calcula, pero te voy a enseñar una **técnica especial** para que lo averigües tú mismo.

Imagina que el helado está en contenedores, podrías decir "sáltate el primero, después 3 paladas, después sáltate los 3 contenedores siguientes" ¡y acabarás con 3 paladas de chocolate!

b	c	l	s	v

Entonces es como si ordenaras a un robot que te trajera helado, pero no cambia nada, tendrás lo que quieres.

Ahora puedes escribirlo como →○○○→→→(la flecha es saltar, el círculo es tomar)

Entonces los tres ejemplos de arriba se pueden escribir así:

{c, c, c} (3 de chocolate):	→○○○→→→
{b, l, v} (uno de banana, uno de limón y uno de vainilla):	○→→○→→○
{b, v, v} (uno de banana, dos de vainilla):	○→→→→○○

OK, entonces ya no nos tenemos que preocupar por diferentes sabores, ahora tenemos un problema ***más simple*** para resolver: "de cuántas maneras puedes ordenar flechas y círculos"

Fíjate en que siempre hay 3 círculos (3 paladas de helado) y 4 flechas (tenemos que movernos 4 veces para ir del contenedor 1º al 5º).

Así que (en general) hay ***r + (n−1)*** posiciones, y queremos que ***r*** de ellas tengan círculos.

Esto es como decir "tenemos ***r + (n−1)*** bolas de billar y queremos elegir ***r*** de ellas". Es decir, es como el problema de elegir bolas de billar, pero con números un poco distintos. Lo podrías escribir así:

$$\binom{n+r-1}{r} = \frac{(n+r-1)!}{r!(n-1)!}$$

donde ***n*** es el número de cosas que puedes elegir, y eliges ***r*** de ellas
(Se puede repetir, el orden no importa)

Es interesante pensar que podríamos habernos fijado en flechas en vez de círculos, y entonces habríamos dicho "tenemos ***r + (n−1)*** posiciones y queremos que ***(n−1)*** tengan flechas", y la respuesta sería la misma...

$$\binom{n+r-1}{r} = \binom{n+r-1}{n-1} = \frac{(n+r-1)!}{r!(n-1)!}$$

¿Qué pasa con nuestro ejemplo, cuál es la respuesta?

(5+3−1)!	=	**7!**	=	**5040**	**= 35**
3!(5−1)!		**3!×4!**		**6×24**	

TEOREMA DE BAYES.

En su forma algebraica más simple, el teorema de Bayes se refiere al cálculo de la probabilidad condicional del evento A, dado que ha ocurrido el evento B, la forma general del teorema de Bayes es:

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

La fórmula anterior es simplemente una forma específica de la fórmula general para la probabilidad condicional; sin embargo, la importancia especial del teorema de Bayes consiste en que se aplica en el contexto de eventos secuenciales y, además, en que la versión de cálculo de la fórmula proporciona la base para determinar la probabilidad condicional de un evento que ha ocurrido en la primera posición secuencial, dado que se ha observado un evento específico en la segunda posición secuencial. La forma de cálculo para el teorema específico de Bayes es:

$$P(A/B) = \frac{P(A)\, P(B/A)}{P(A1)\, P(B/A1) + P(A2)\, P (B/A2)4... + P(An)\, P(B/An)}$$

Ejemplo 1. La urna A1 contiene 7 bolas blancas y 3 negras; la urna A2 contiene 4 bolas blancas y 9 negras y la urna A3 contiene 6 blancas y 4 negras. Se lanza un dado no cargado. Si resulta 1,2 o 3, se saca una bolita de urna A1; si resulta 4 o 5, la bolita se saca de la urna A2 y finalmente si resulta 6 se saca de la urna A3. Dado que la bolita extraída fue blanca. ¿Cuál es la probabilidad de que ella provenga de la urna A2?

$$P(A2/B) = \frac{P(A2)\, P(B/A2)}{P(A1)\, P(B/A1) + P(A2)\, P(B/A2) + P(A3)\, P(B/A3)}$$

$$P(A2/B) = \frac{(2/6)\,(4/13)}{(3/6)(7/10) + (2/6)(4/13) + (1/6)(6/10)}$$

$$P(A2/B) = \frac{(8/78)}{(21/60) + (8/78) + (6/60)} = \frac{0.1025}{0.35 + 0.10 + 0.1} = \frac{0.1025}{0.55} = \mathbf{\underline{0.18}}$$

Ejemplo 2. Tres máquinas A, B y C producen respectivamente 50%, 30% y 20% del número total de refrescos de una fábrica, embotelladora. Los porcentajes de desperfectos de producción de estas máquinas son 4%, 5% y 6%, si se selecciona al azar un artículo, hallar la probabilidad de que el articulo sea defectuoso.

X= Defectuosos

$$P(X)= P(A)\ P(X/A) + P(B)\ P(X/B) + P(X)\ P(X/C)$$

$$P(X)= (0.50)\ (0.04) + (0.30)\ (0.05) + (0.20)\ (0.06)$$

$$P(X)= 0.02 + 0.015 + 0.012$$

$$P(X)= 0.047$$

Ejemplo 3. Si del ejemplo anterior, se selecciona un refresco y es defectuoso, hallar la probabilidad de que el artículo fue producido la maquina A.

$$P(A/X)= \frac{P(A)\ P(X/A)}{P(A)\ P(X/A) + P(B)\ P(X/B) + P(C)\ P(X/C)}$$

$$P(A/X)= \frac{(0.50)\ (0.04)}{(0.50)\ (0.04) + (0.30)\ (0.05) + (0.20)\ (0.06)} = \frac{0.02}{0.02 + 0.015 + 0.012} = \frac{0.02}{0.04} = \underline{\mathbf{0.42}}$$

Ejercitando mi habilidad. (Producto esperado)

Problema 1: En tu cuaderno diseña el diagrama de árbol correspondiente a la situación donde se tiene una urna que contiene 3 pelotas con las letras A, B y C, otra urna con 4 pelotas con los números del 1 al 4. Si se elige una pelota de cada urna ¿Cuántas parejas distintas se pueden formar?

Problema 2: En tu cuaderno diseña el diagrama de árbol correspondiente al problema donde un médico general clasifica a sus pacientes de acuerdo a su sexo (masculino o femenino), tipos de sangre (A, B, AB u O) y en cuanto a la presión sanguínea (normal, alta y baja), ¿Cuántas clasificaciones tiene el médico?

Problema 3: Dados los conjuntos: A={0,1,2,3,4,5}, B={3,5,7} y C={2,4}, en tu cuaderno resuelve las operaciones:

a) $A \cap B$ b) $B \cup C$ c) $B - A$ d) C^c

Problema 4: Dados los conjuntos A={0,1,2,3,4,5}, B={0,2,4} y C={5,6,8}, resuelve las operaciones:

a) $B \cup C$

b) $A \cap C$

Problema 5: Calcular cuantas placas para automóvil es posible formar para la Ciudad de México, si tienen 3 letras y 3 dígitos y la primera letra solo es A o B considerando un alfabeto de 26 letras.

Problema 6: ¿Cuántos comités diferentes se pueden formar con 40 alumnos, si cada comité está formado por un presidente, un secretario, un tesorero y 2 vocales?

Problema 7: ¿Cuántas alternativas tiene un estudiante para seleccionar su carrera, si la facultad de Ciencias de la UNAM le ofrece 3 carreras, en la de Ingeniería le ofrecen 5 y en Ciencias Biológicas le ofrecen 6, considerando que no le interesa otra facultad?

Problema 8: ¿De cuantas formas diferentes puede Samy elegir su educación preparatoria entre 3 escuelas que le ofrecen Bachillerato General por competencias, y 2 que le ofrecen Bachillerato Técnico o Terminal, si solo puede elegir uno de ellos?

Problema 9: ¿De cuantas formas se puede pintar una bandera de 3 franjas, si se dispone de 8 colores?

Problema 10: ¿Cuántas permutaciones se pueden obtener con todas las letras de la palabra “tenis”?

Problema 11: ¿De cuantas formas diferentes se pueden acomodar 9 personas en una camioneta que tiene 3 asientos disponibles?

Problema 12: En una caja hay 8 fichas, marcadas con las letras A, B, C, D, E, F, G y H. Se van a tomar 6 fichas al azar. ¿De cuantas formas diferentes se pueden tomar las fichas?

¿Qué Aprendí?

Problema 1: Para el caso de que una máquina empiece a trabajar con una clave que se conforma con los dígitos de los botones, por ejemplo; AAA, BBB, CCC, AAB, …

Construye todas las soluciones posibles.

Problema 2: Utilizando las letras de la palabra AMOR, construye en tu cuaderno todas las palabras con las siguientes características:

a) Con significado, ejemplo; ROMA.

b) Sin significado, ejemplo; AORM.

Problema 3: Supongamos que una empresa clasifica a sus empleados de la siguiente manera: Existen 14 empleados con buen sueldo (S), 11 empleados con un buen futuro (F), 14 trabajadores con un buen plan de jubilación (J), además se conoce que la compañía tiene 3 empleados con buen sueldo y con buen futuro, 5 con buen sueldo y buen plan de jubilación, 4 subordinados con buen futuro y buen plan de jubilación y un dependiente con buen sueldo, futuro y plan de jubilación.

a) ¿Cuántos empleados tiene la compañía?

b) ¿Cuántos empleados tienen solo buen suelo?

c) ¿Cuántos empleados tienen solo buen futuro?

d) ¿Cuántos empleados tienen solo buen plan de jubilación?

Problema 4: Sean U = {1,2,3,4,5,6,7,8,9}, A = {2,4,6}, B = {1,3,5}, C = {7,8,9}, realizar las operaciones:

a) $A \cup B$

b) $A - C$

c) $A \cap B$

d) A^C

Rescatando mis Aprendizajes

Se le pide amablemente consulte el links www.geogebra.org/t/probability donde se encontrará con un mapa conceptual, con el tópico de Probabilidad al centro y ramificaciones a la orilla con los conceptos como Diagrama de árbol, Condicional, Experimento aleatorio y Combinatoria. Usted realizara lo siguiente:

1) Cuando usted selecciona el tema de Probabilidad, en la parte de abajo de la pantalla, aparecen una serie de actividades y recursos, usted seleccionará de entre esas actividades de 4 a 6 de ellas e interactuará con los recursos, manipulando las opciones y registrará esos ejercicios como conclusión al respecto.

2) Posteriormente seleccionará el tema de Combinatoria, buscando de entre 2 y 4 actividades de su interés en la parte inferior de la pantalla, interactuando con los recursos y registrando los ejemplos.

3) Enseguida seleccione el tema de Experimentos aleatorios y busque 4 actividades que le llamen la atención en la parte inferior de la pantalla, interactúe con ellos y registre los ejemplos al respecto.

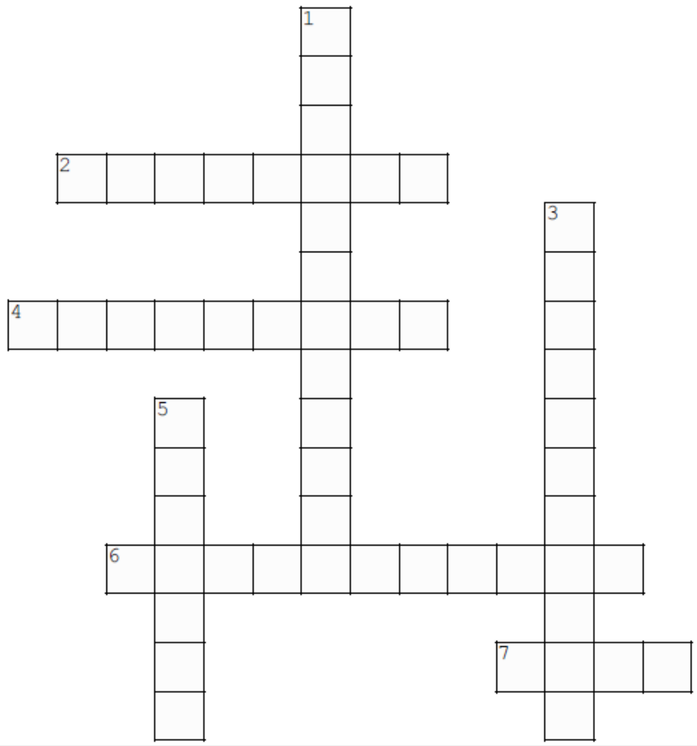
4) Así mismo se le pide seleccionar el tema Condicional, y elegir de 2 a 4 actividades en la parte inferior de la pantalla, interactuar con ellos y registrar los ejemplos.

5) Finalmente selecciona el tema Diagrama de árbol, elige 2 actividades en la parte inferior y de igual manera interactúa con ellos, registrando los ejemplos como evidencia de análisis.

CONCEPTOS ESTADÍSTICOS

Nombre: ____.

Resuelve el siguiente crucigrama.



Horizontal

2. Característica de interés de un elemento.

4. Colección de objetos

6. Ciencia que permite recolectar, organizar, procesar, analizar e interpretar datos.

7. Conjunto de valores de la variable medida.

Vertical

1. La posibilidad que ocurra un evento sin saber su resultado.

3. Efectúan estimaciones, decisiones, predicciones sobre un conjunto mayor de datos.

5. Subconjunto de elementos

Actividad Transversal

El docente te dará indicaciones de la actividad trasversal que deberás realizar.

UNIDAD II

UNIDAD II.

INTRODUCCIÓN

La estadística descriptiva nos permite proporcionar métodos para organizar y resumir datos obtenidos por medio de un experimento estadístico, pueden ser por medio de gráficos o tablas, y así expresarlos en una interfaz más amigable, es decir; será más sencillo acceder a la información a través de un gráfico que analizando una población o muestra de datos. Estos métodos sirven para tener un mejor uso de la información y también permite obtener conclusiones y así tener una idea más precisa de las características de los datos.

La mayor parte de las veces, es poco probable que se puedan utilizar todos los datos de una población, pero si se toma una muestra representativa, es más sencillo estudiarla, realizar inferencias y describir el conjunto total de datos.

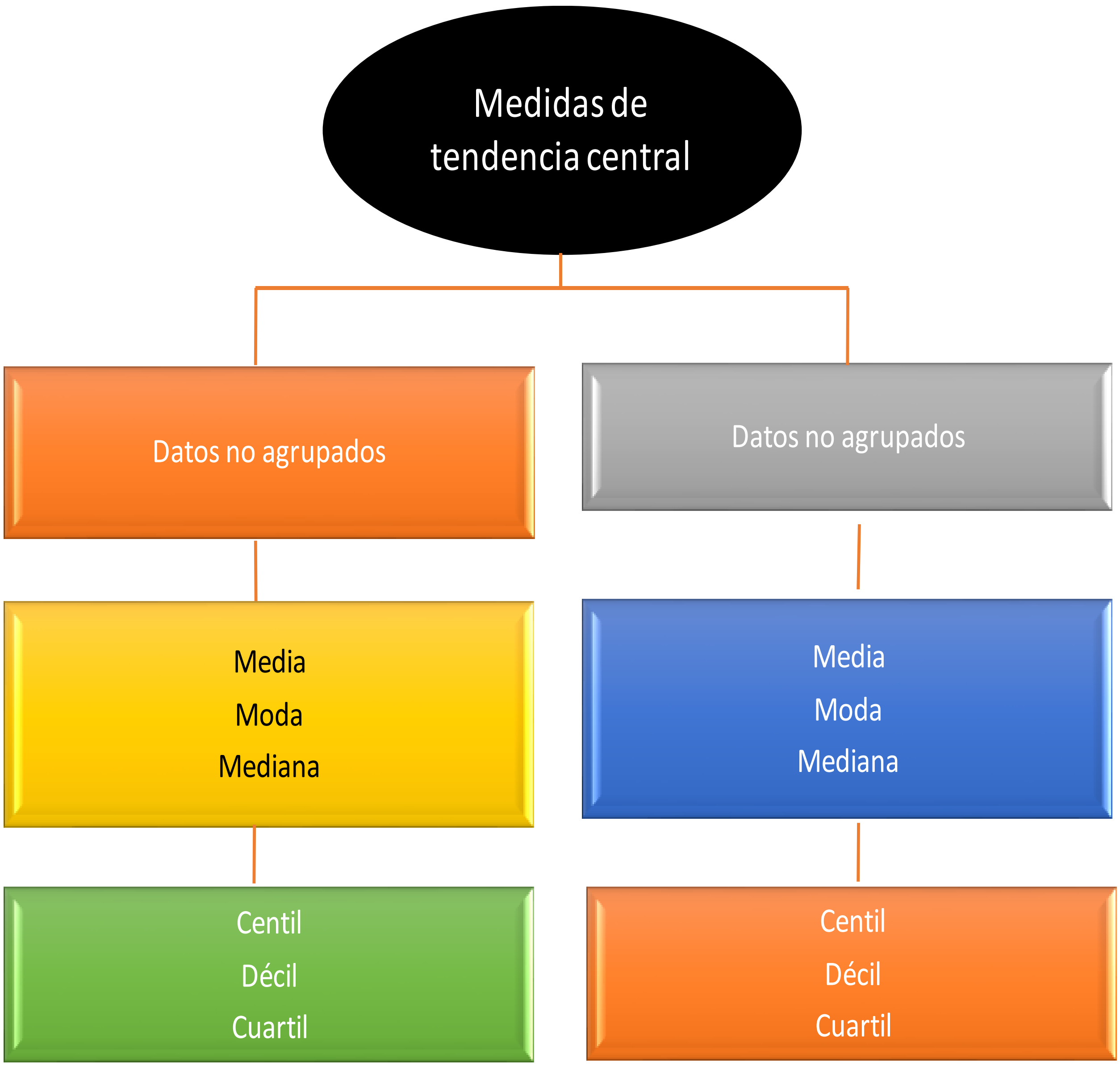
En esta unidad vamos a analizar las tablas de frecuencias y sus componentes, así como la construcción de gráficos como el Histograma, Polígono de Frecuencias, Grafica de Barras y Graficas de Porcentaje (pastel) entre otras.

Luego de desarrollar un experimento estadístico es necesario ordenar la información, para ello se puede hacer de forma ascendente o descendente, esto es, que se puede realizar comenzando desde el menor número hasta el mayor (ascendente) o del mayor número al menor (descendente), es imprescindible que cuando trabajemos con algún ejercicio en Estadística ordenar la muestra de datos.
En esta unidad veremos los temas de:

- Medidas de tendencias central
- Diagramas de frecuencia
- Gráficos estadísticos

CONTENIDO

BLOQUE 2



1. MEDIDAS DE TENDENCIA CENTRAL

Estudiarás que las medidas de tendencia central que junto con otras medidas se utilizan para describir y establecer comparaciones cuantitativas entre distribuciones de frecuencia.

Lo que debes traer en tu mochila cerebral

TEMA	DÓNDE LO PUEDES REPASAR
Manejo de operaciones básicas de aritmética	Matemáticas Secundaria
Habilidad manual para el uso de calculadora	Continuo
Conceptos de conjuntos	Álgebra I
Lectura de comprensión	Español I

Saberes para desarrollar las competencias integradas

Saber conocer	**Saber hacer**	**Saber ser**
Conocimientos • Reconoce la importancia de las medidas de la tendencia central. • Distingue las medidas para datos agrupados y No agrupados.	**H**abilidades • Calcula las medidas de tendencia central. • Manejo eficiente de calculadora u otro equipo de cálculo. • Aplica conocimientos básicos de aritmética.	**A**ctitudes y valores • Desarrolla una actitud crítica y reflexiva sobre las medidas de tendencia central. • Asume una actitud crítica y responsable en el trabajo en equipo.

Para aprender más

Ejemplo: se tiene los siguientes datos para el peso de hombres, en kilogramos, que acudieron a la clínica de prevención del IMSS:

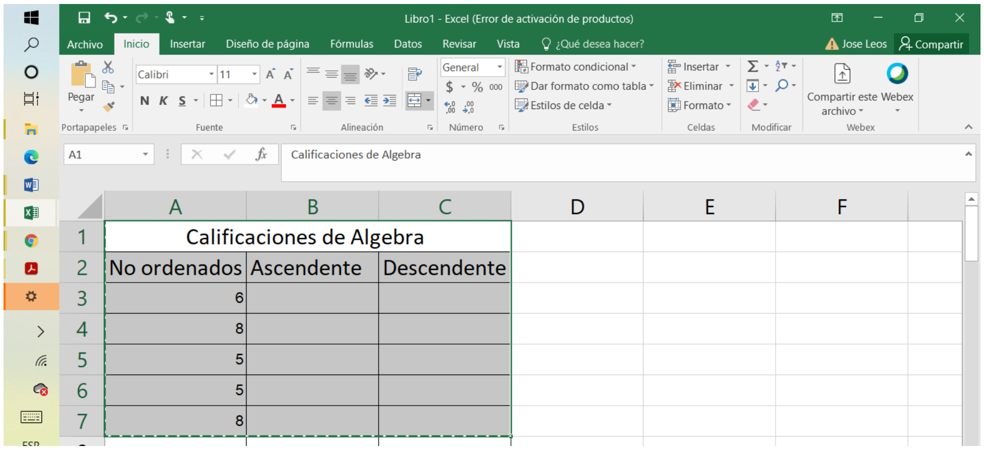
Datos de peso	75	82	77	76	80	76	82	73	72	83
Ordenados ascendente	72	73	75	76	76	77	80	82	82	83
Ordenados descendente	83	82	82	80	77	76	76	75	73	72

Si observamos se escriben todos los datos de la información proporcionada por el ejercicio independientemente si se repiten o no, la primera fila nos muestra los datos tal cual nos los proporciona el ejercicio, la segunda fila los datos ordenados de menor a mayor (ascendente) y la tercera los datos ordenados de mayor a menor (descendente).

Esto se puede hacer de forma manual o también utilizando programas que se tienen en cualquier computadora, a continuación, se muestra cómo usar la opción ordenar en Excel.

Primero se transcriben todos los datos del ejercicio a una hoja de cálculo de Excel o algún programa donde se puedan generar hojas de cálculo. Se hace un encabezado que especifique los datos que se van a tratar, en este caso: “Calificaciones de Álgebra”, Se combinaron y centraron las tres primeras celdas y luego en forma vertical se agregaron las calificaciones del grupo, la primera columna se llama “No ordenados” la segunda “Ascendente” y la tercera “Descendente”.

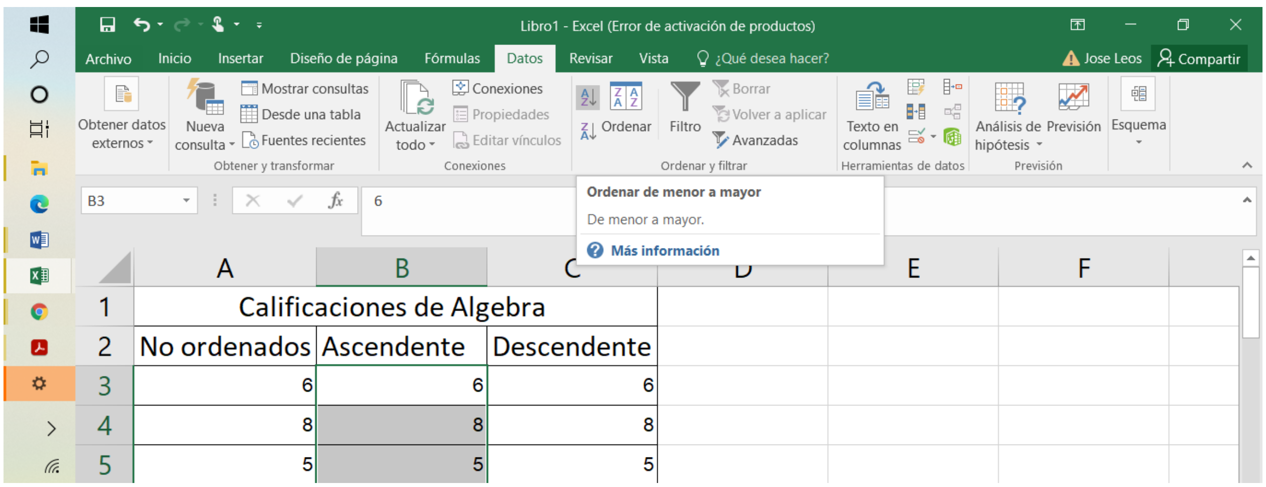
Imagen No. 3 Tendencias



Fuente: Elaboración propia, noviembre 2020

Se copian los datos a cada una de las columnas, se selecciona los datos de la columna “Ascendente”, se da clic a la pestaña datos y se selecciona el botón ordenar A→Z y del menú de iconos.

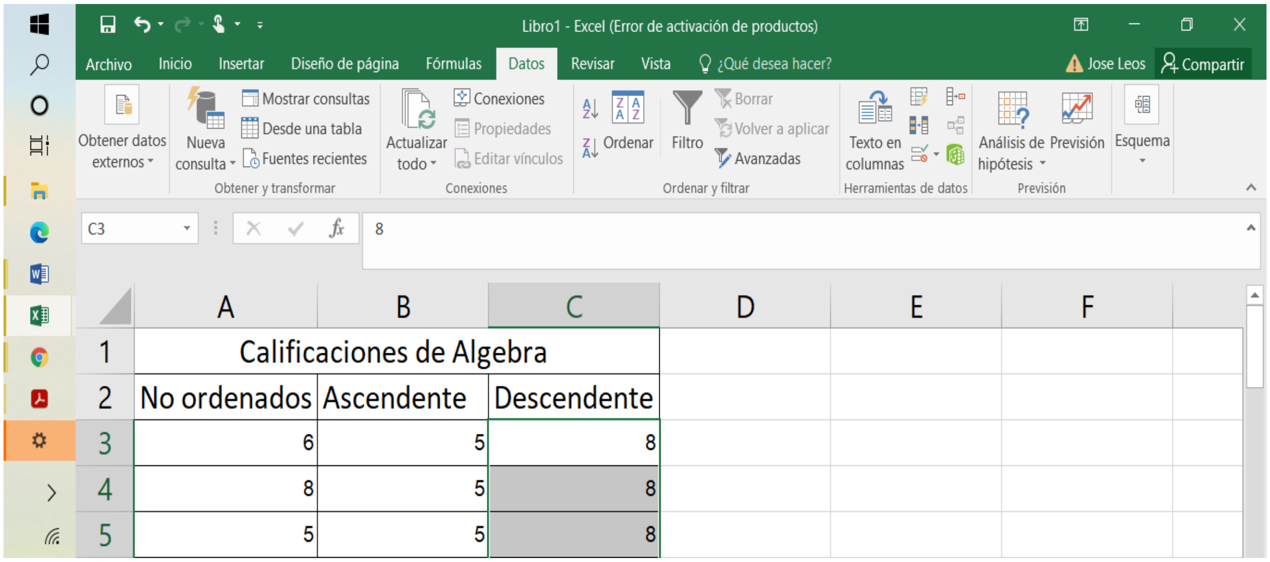
Imagen No. 4 Ejemplos



Fuente: Elaboración propia, noviembre 2020

Se hace lo mismo con la Columna “Descendente” pero ahora se selecciona el icono ordenar Z→A, el resultado es el siguiente:

Imagen No. 5 Ejemplo



Fuente: Elaboración propia, noviembre 2020

Si observamos la columna “Ascendente” empieza con el 5 y la columna “Descendente” empieza con el 8.

Ejercitando mi habilidad.

Instrucciones. - Ordena los siguientes datos de forma manual y también utilizando una hoja de cálculo.

Ejemplo: Los siguientes datos corresponden a las estaturas de los alumnos de 4to semestre del grupo 401 de la carrera de Instrumentación.

1.75	1.82	1.77	1.76	1.80	1.76	1.82	1.73	1.72	1.83
1.72	1.73	1.75	1.76	1.76	1.70	1.80	1.60	1.82	1.83
1.83	1.82	1.82	1.62	1.77	1.76	1.76	1.65	1.73	1.72
1.65	1.67	1.70	1.79	1.78	1.65	1.67	1.60	1.66	1.82

Puedes hacerlo de forma manual en el siguiente formato de tabla de limpio:

1.60									
									1.83

O puedes hacerlo en el programa de tu elección que genere hojas de cálculo.

Para aprender más

TABLAS DE FRECUENCIA.

Las tablas de frecuencia son instrumentos matemáticos que evitan cálculos tediosos, ya que para datos que muestran un amplio rango, estos se pueden agrupar en clases o valores, a continuación, se definirá cada uno de los datos necesarios para calcular las tablas de frecuencia, así como su nomenclatura.

Las tablas de frecuencia se construyen a partir de ciertos parámetros, los cuales definiremos a continuación:

Frecuencia absoluta (f_i) La frecuencia absoluta es el número de veces que aparece un determinado valor en un estudio estadístico. La suma de las frecuencias es igual al número total de datos, que se representa por N.

Frecuencia acumulada (F_i). La frecuencia acumulada es la suma de las frecuencias absolutas de todos los valores inferiores o iguales al valor considerado.

Frecuencia relativa (f_r). La frecuencia relativa es el cociente entre la frecuencia absoluta de un determinado valor y el número total de datos. La frecuencia es igual a la unidad.

Para datos no agrupados como por ejemplo las calificaciones de los alumnos, al ser números consecutivos y, se consideran como datos no agrupados ya que no hay un gran rango de valores y las variables son no continuas, para construir la tabla se hace lo siguiente:

Primero debemos tener una base de datos.

Calificaciones Álgebra									
6	6	6	6	6	6	6	6	6	6
8	8	8	8	8	8	8	8	8	8
5	5	5	6	5	6	5	6	5	5
6	5	7	5	7	5	7	7	7	6

Se ordenan de formas ascendente.

Calificaciones Álgebra									
5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6
6	6	6	6	6	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8

Se cuenta el número de veces que un dato se repite y se estructura la tabla de frecuencias para datos no agrupados como sigue:

$$f_R = \frac{f_i}{N}$$

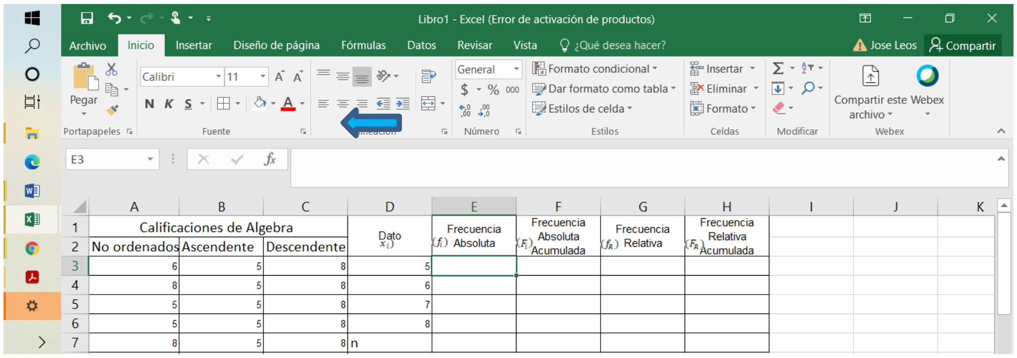
Dato (x_i)	Frecuencia Absoluta (f_i)	Frecuencia Absoluta Acumulada (F_i)	Frecuencia Relativa (f_R)	Frecuencia Relativa Acumulada (F_R)
5	10	10	0.25	0.25
6	15	25	0.375	0.625
7	5	30	0.125	0.75
8	10	40	0.25	1.0
N	40	$\sum f_R$	1.0	

De acuerdo a las definiciones antes mencionadas la frecuencia absoluta corresponde al número de veces que se repite un dato, por ejemplo: el 5 se repite 10 veces, y el 7 se repite 5 veces. La frecuencia absoluta acumulada parte del primer dato de frecuencia absoluta y se van agregando las siguientes frecuencias absolutas.

Si observan aparece el parámetro (n) que corresponde al número total de datos en la muestra o población. Y el igual al último número que aparece en la columna de la frecuencia acumulada, esto puede servir para ir comprobando que nuestro ejercicio está bien realizado. La frecuencia relativa corresponde (f_R) al cociente o la división entre la frecuencia absoluta y el número total de datos en la muestra o población, la sumatoria de las frecuencias relativas $(\sum f_R)$ es siempre 1.0, la frecuencia relativa acumulada parte del primer dato de frecuencia realtiva y se van agregando las siguientes frecuencias realtivas. El último valor de la frecuencia relativa acumulada debe ser 1.0.

Las tablas de frecuencia también pueden realizarse en Excel para ello vamos a retomar el archivo que previamente generamos y se utilizaran funciones como: “contar sí”, “división” y “suma” y “autocompletar”. Lo primero es poner el encabezado y capturar la columna Dato.

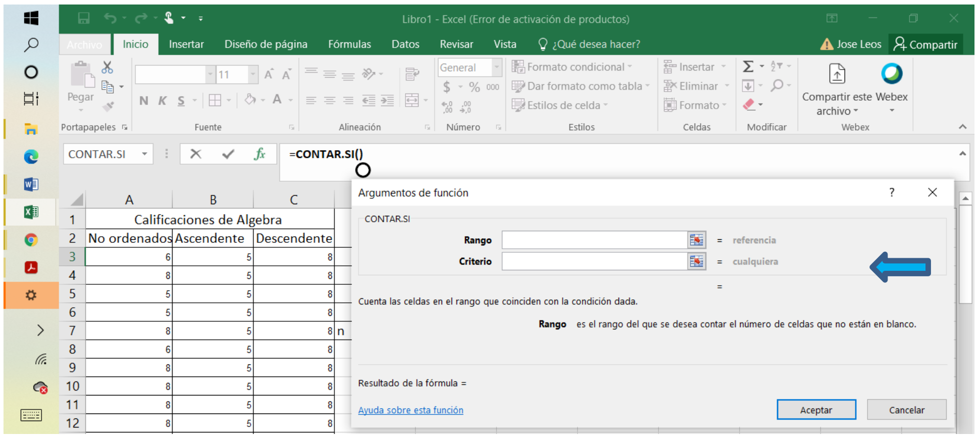
Imagen No. 6 Ejercicios



Fuente: Elaboración propia, noviembre 2020

Para calcular la columna de la frecuencia absoluta se utiliza la función “contar.si” en el menú de opciones se selecciona el icono (f_x) ese botón activa todo el catálogo de fórmulas que se pueden usar en la hoja de cálculo. En la opción buscar se escribe “contar.si” y cuando aparezca en el submenú se selecciona la función, se le dará aceptar y aparecerá una ventana emergente con unos submenús que dicen “rango” y “criterio” como se muestra a continuación.

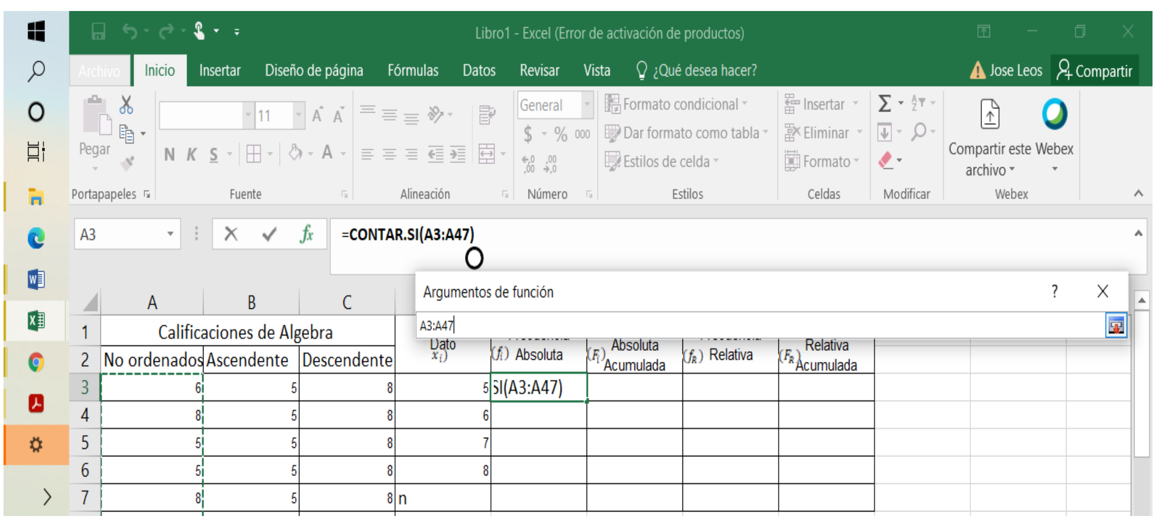
Imagen No. 7 Ejercicio



Fuente: Elaboración propia, noviembre 2020

Donde aparece rango se seleccionarán todas las celdas donde están los datos que ya fueron ordenados de forma ascendente y en el criterio se selecciona el número 5 de la celda de datos. Para ello se da clic sobre las flechas rojas a un lado del recuadro que sigue a la palabra rango o criterio.

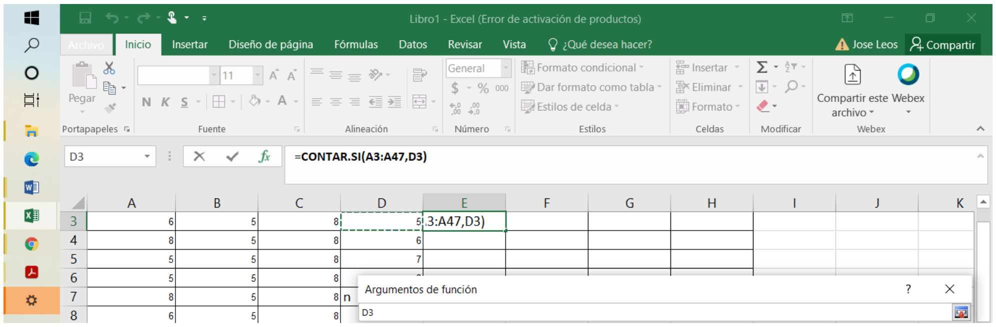
Imagen No. 8 Ejercicio



Fuente: Elaboración propia, noviembre 2020

Al dar clic de nuevo sobre la flecha roja al lado del recuadro se guardan esos datos y posteriormente se selecciona el criterio a contar, que en este caso es 5.

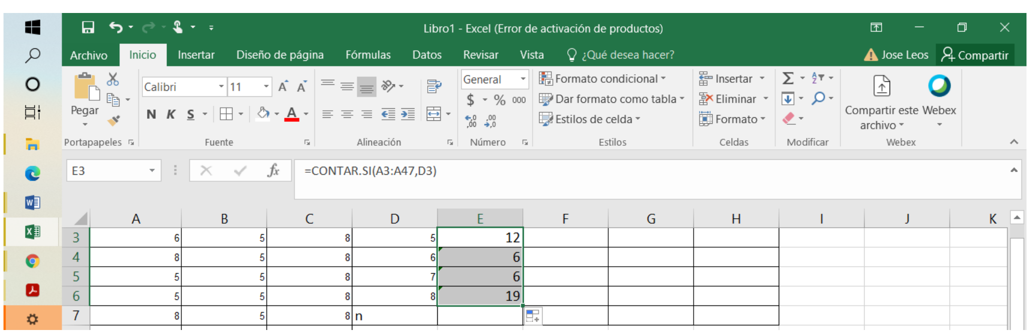
Imagen No. 9 Ejercicios



Fuente: Elaboración propia, noviembre 2020

Se da de nuevo clic a la fecha roja y luego en aceptar y ya nos cuenta el número de veces que se repite el 5 en nuestra muestra de datos. A continuación, se utiliza la función “autocompletar” que sirve para repetir la misma operación que se hizo anteriormente sin tener que llevar a cabo todos los pasos que se realizaron con anterioridad y nos queda la columna de la frecuencia absoluta completa. Para activar la función autocompletar es necesario ubicar el puntero del mouse sobre la esquino inferior derecha hasta que aparezca una cruz negra y posteriormente arrastrar sobre las celdas que se necesitan autocompletar.

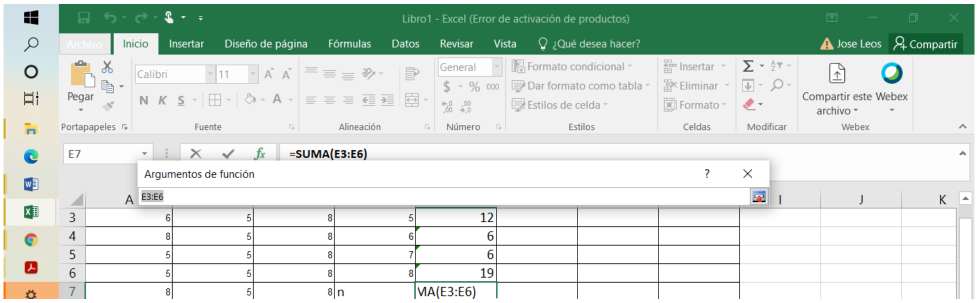
Imagen No. 10 Ejercicio



Fuente: Elaboración propia, noviembre 2020

Para calcular (N) es necesario usar la función suma del catálogo de funciones que se activan desde el recuadro de (f_x), para ello nos ubicamos en la parte inferior de la columna de la frecuencia absoluta y en esa celda activamos la función suma, automáticamente se seleccionarán los datos numéricos de las celdas anteriores y al dar enter se obtendrá el resultado de la suma, que equivale al total de datos de la muestra.

Imagen No. 11 Ejercicios

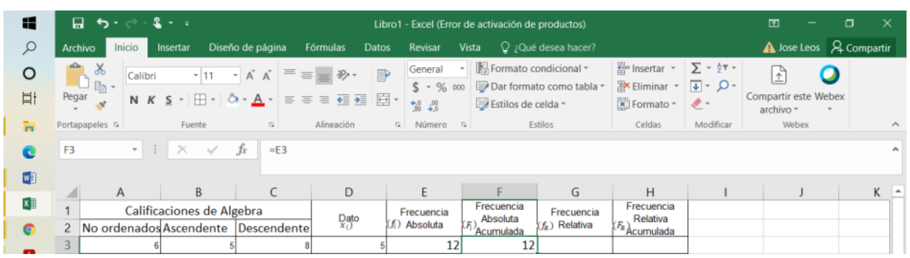


Fuente: Elaboración propia, noviembre 2020

La siguiente columna se calcula situándonos sobre la celda debajo del título de frecuencia absoluta acumulada y se realizan los siguientes pasos:

1. Se duplica el primer valor de la frecuencia absoluta usando el operador igual "=" y luego se selecciona la celda correspondiente y se teclea enter.

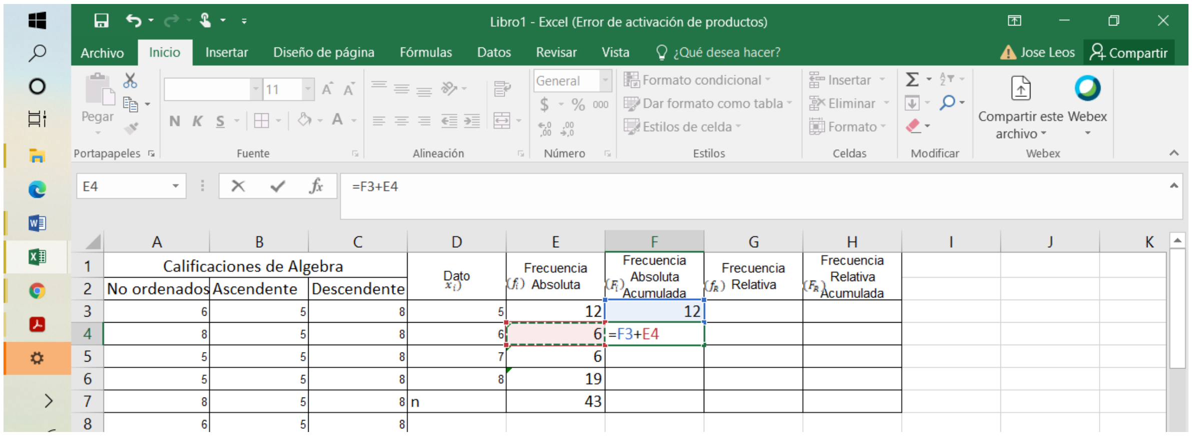
Imagen 12 Ejercicio



Fuente: Elaboración propia, noviembre 2020

2. En la siguiente celda sobre la columna se vuelve a usar el operador igual "=" se selecciona la celda superior, luego se utiliza el operador suma "+" y se selecciona la celda de la izquierda y se teclea enter.

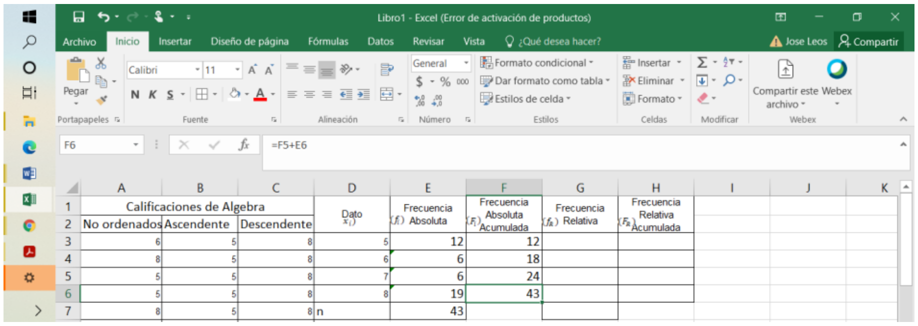
Imagen No. 13 Ejercicio



Fuente: Elaboración propia, noviembre 2020

3. Por último, se usa la función autocompletar y se completan las demás celdas.

Imagen No 14 Ejercicio

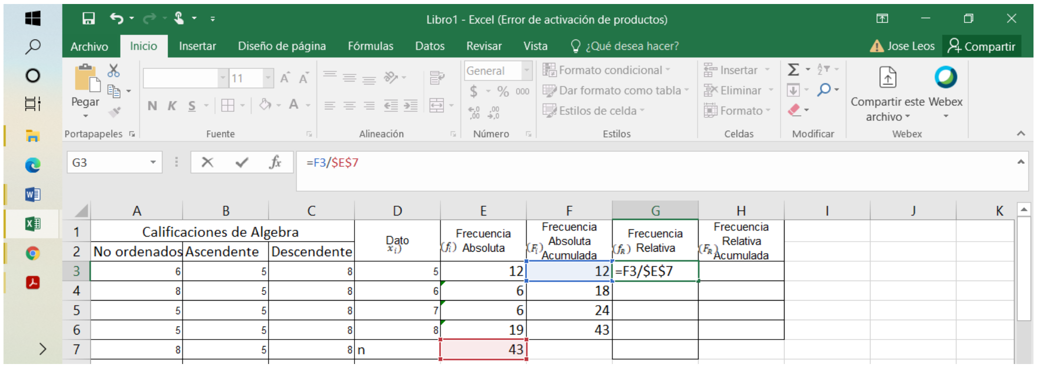


Fuente: Elaboración propia, noviembre 2020

Para calcular la frecuencia relativa debemos dividir el valor de la frecuencia que corresponda a cada dato entre (N) que es el número total de datos. Al momento de dividir el valor de (N) lo haremos con la celda donde se encuentra su valor y para que siempre se mantenga el mismo valor debemos anteponer un signo de peso ($) antes de la letra y antes del número que identifica a las celda.

En este caso corresponde a la celda E7 por lo que tendríamos $E$7. Para hacer la operación se escribe le operador "=" se selecciona la celda de la frecuencia correspondiente se escribe el operador "/" luego $E$7 y se teclea enter.

Imagen No. 15 Ejercicio



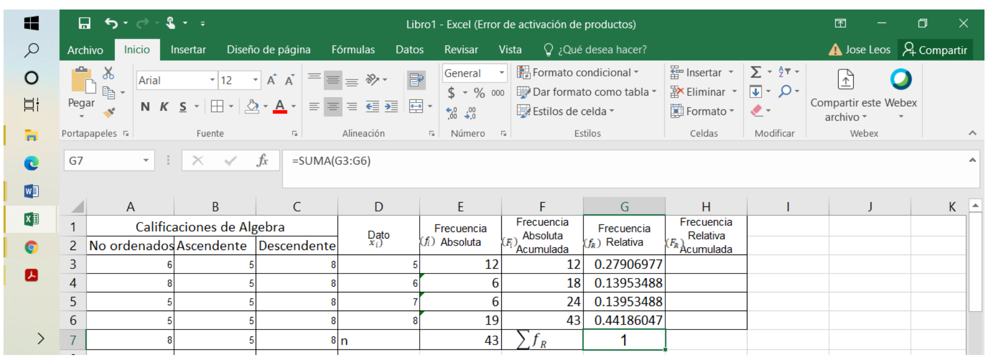
G3: =F3/$E$7

Calificaciones de Algebra			Dato x_i	Frecuencia Absoluta (f_i)	Frecuencia Absoluta Acumulada (F_i)	Frecuencia Relativa (f_R)	Frecuencia Relativa Acumulada (F_R)
No ordenados	Ascendente	Descendente					
6	5	8	5	12	12	=F3/$E$7	
8	5	8	6	6	18		
5	5	8	7	6	24		
5	5	8	8	19	43		
8	5	8	n	43			

Fuente: Elaboración propia, noviembre 2020

Al finalizar se utiliza la función “autocompletar” y se completan la columna de la frecuencia relativa.

Imagen No. 16 Ejercicio



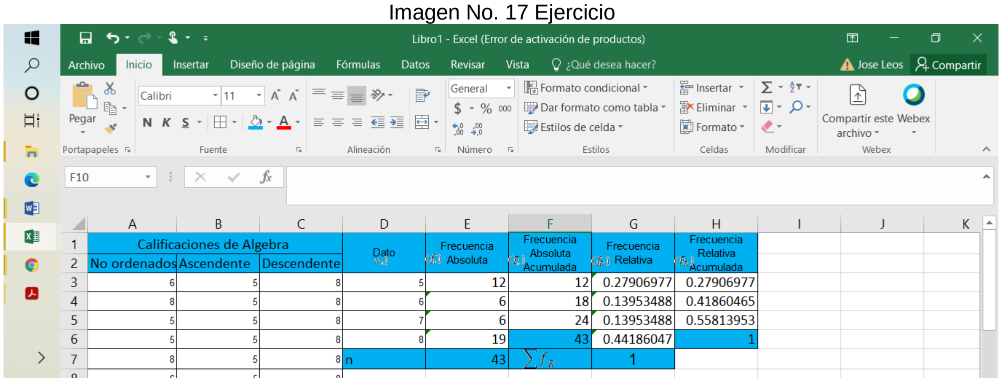
G7: =SUMA(G3:G6)

Calificaciones de Algebra			Dato x_i	Frecuencia Absoluta (f_i)	Frecuencia Absoluta Acumulada (F_i)	Frecuencia Relativa (f_R)	Frecuencia Relativa Acumulada (F_R)
No ordenados	Ascendente	Descendente					
6	5	8	5	12	12	0.27906977	
8	5	8	6	6	18	0.13953488	
5	5	8	7	6	24	0.13953488	
5	5	8	8	19	43	0.44186047	
8	5	8	n	43	$\sum f_R$	1	

Fuente: Elaboración propia, noviembre 2020

Revisa como se calculó la Frecuencia Absoluta Acumulada y repite los pasos al final la tabla de frecuencias queda como sigue:

Imagen No. 17 Ejercicio



Calificaciones de Algebra			Dato x_i	Frecuencia Absoluta (f_i)	Frecuencia Absoluta Acumulada (F_i)	Frecuencia Relativa (f_R)	Frecuencia Relativa Acumulada (F_R)
No ordenados	Ascendente	Descendente					
6	5	8	5	12	12	0.27906977	0.27906977
8	5	8	6	6	18	0.13953488	0.41860465
5	5	8	7	6	24	0.13953488	0.55813953
5	5	8	8	19	43	0.44186047	1
8	5	8	n	43	$\sum f_R$	1	

Fuente: Elaboración propia, noviembre 2020

1.2 Medidas de tendencia central para **datos No agrupados.**

En las medidas de tendencia central para datos no agrupados la frecuencia no se toma en cuenta, mientras que las medidas de tendencia central para datos agrupados si se toma en cuenta la frecuencia de las mismas.

1.2.1 Media aritmética

La media aritmética de cierta cantidad de datos es la suma de sus valores divididos por el número de ellos, matemáticamente la podemos expresar de la siguiente forma:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_m}{N}$$

Donde:
N = Número de datos
X_m: Valor que toma la variable o cada dato.
$\bar{X}$: Media aritmética

1.2.2 Moda

Es una distribución de frecuencias la moda es el dato que más veces se repite o es el valor más frecuente en un conjunto de datos. En ocasiones se presentan dos o más valores que se repitan con mayor frecuencia y si eso ocurre a los datos se les conoce como bimodal o multimodal.

La moda de los datos la podemos obtener ordenando de menor a mayor, de la tabla de puntos.

Observa qué datos es el de mayor frecuencia, por lo tanto, la moda se declarará:

m_o = Moda

1.2.3 Mediana

La mediana es el valor del elemento de la posición central de los datos individuales, ordenadas de menor a mayor o viceversa, y es el punto que marca la mitad de los valores mayores que él y la mitad de los valores menores que él, es decir, está a la mitad, con el 50% de los valores a su derecha y el 50% de los valores a su izquierda.

Para calcular la mediana los datos los datos tienen que estar ordenados como se mencionó anteriormente, luego se calcula la posición de la misma bajo el siguiente criterio:

- Si se tiene un número impar de datos ordenados y la mediana será única.
- Si el número de datos es par no hay término del medio, en este caso la mediana será la suma de los datos al centro, luego se suman y se dividen entre dos.

La mediana ocupa la posición central y se declara:

m_e = mediana

Imagen No. 18 Ilustración de mediana

Fuente: https://www.pinterest.es/pin/9640586683876305/ Recuperado 23 de noviembre 2020

Toma de datos. Consiste en recopilar datos de una población que no han sido organizados numéricamente.

Ordenación. Es un conjunto de datos numéricos en orden creciente o decreciente.

Distribución de frecuencias. Es una disposición tabular de los datos por clases junto con las correspondientes frecuencias de clase. Es una agrupación de datos en categorías mutuamente excluyentes (es un individuo, objeto, medición que pertenece únicamente a una categoría) dando el número de observaciones en cada categoría.

Clase o categoría. Es un intervalo en que pueden agruparse los datos de una variable estadística.

Intervalos de clase. Son cada uno de los intervalos en que pueden agruparse los datos, es decir, es el símbolo que define una clase. Con frecuencia se intercambian los términos clase e intervalo de clase, aunque el intervalo de clase es un símbolo para la clase.

Frecuencias de clase. Es determinar el número de individuos que pertenecen a cada clase.

Límites de clase. Son los números extremos de una clase (intervalo de clase) y se les denomina límite inferior de clase y límite superior de clase.

Limites reales de clase (Fronteras de clase). Son las fronteras de clase que se obtienen promediando el límite superior de una clase con el límite inferior de la siguiente. En estos el menor es la frontera inferior y el mayor la frontera superior. A veces se usan las fronteras de clase como símbolos para la clase, sin embargo, las fronteras no deben coincidir con valores realmente medidos.

Tamaño del intervalo de clase (anchura del intervalo de clase). Es la diferencia entre las fronteras de clase superior e inferior. Si todos los intervalos de clase de una distribución de frecuencias tienen la misma anchura, la denotaremos por C.

Marca de clase. Es el punto medio del intervalo de clase y se obtiene promediando los limites inferior y superior de clase (punto medio de la clase).

Para construir una tabla de frecuencias para datos agrupados, es necesario primero calcular unos parámetros que nos permitirán delimitar las clases el y número de las mismas, por ejemplo, si retomamos el ejercicio de las estaturas podemos ver que las variables son continuas y tienen decimales por lo que se debe utilizar este tipo de tablas.

1.75	1.82	1.77	1.76	1.80	1.76	1.82	1.73	1.72	1.83
1.72	1.73	1.75	1.76	1.76	1.70	1.80	1.60	1.82	1.83
1.83	1.82	1.82	1.62	1.77	1.76	1.76	1.65	1.73	1.72
1.65	1.67	1.70	1.79	1.78	1.65	1.67	1.60	1.66	1.82

Ya ordenados de forma ascendente los datos de la muestra nos quedarían así:

1.60	1.62	1.62	1.65	1.65	1.65	1.66	1.67	1.67	1.70
1.70	1.72	1.72	1.72	1.73	1.73	1.73	1.75	1.75	1.76
1.76	1.76	1.76	1.76	1.76	1.77	1.77	1.78	1.79	1.80
1.80	1.82	1.82	1.82	1.82	1.82	1.82	1.83	1.83	1.83

A continuación, se definen y se anotan las fórmulas de los parámetros que nos permiten construir nuestra tabla de frecuencias de datos agrupados usando clases.

Rango: es la diferencia entre el dato de mayor valor y el dato de menor valor.

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

$$R = 1.83 - 1.60$$

$$R = 0.23$$

Marca de clase (Xm) La marca de clase es el punto medio de cada intervalo y es el valor que representa a todo el intervalo.

Número de Clase. Son divisiones o categorías en las cuales se agrupan un conjunto de datos ordenados con características comunes. Para organizar los valores de la serie de datos hay que determinar un número de clases que sea conveniente.

Fórmula de Sturges (K): da una idea del número de clases en las cuales dividir nuestros datos.

$$K = 1 + 3.322 \log_{10} N$$

Donde:

K: número de clases

N: número total de datos.

En el ejemplo de las estaturas podemos calcular el número de clases sustituyendo el número total de datos.

$$K = 1 + 3.322 \log_{10}(40)$$
$$K = 6.322 \approx 6$$

El número de clases es redondeable por lo que habrá 6 clases.

Límites de la clase. Cada clase está delimitada por el límite inferior de la clase y el límite superior de la clase, se calcula de acuerdo a la amplitud.

Amplitud de la clase(A). La amplitud de la clase es la diferencia entre el límite superior e inferior de la clase, que se puede calcular con la siguiente fórmula:

$$C = \frac{R}{K} = \frac{0.23}{6} = 0.038 \approx 0.04$$

$$C = 0.038$$

Entonces para empezar a determinar el límite de clase se les sumará 0.038 cada vez que se establezca una clase, empezando por el valor mínimo, es decir, para la primera clase se tiene como límite inferior el 1.60 al que se sumará 0.038 para que el límite superior de la clase sea: 1.638. Al terminar no importa que la última clase no se ajuste a la amplitud, por ejemplo, al sumar al límite inferior de la última clase (1.79) el valor de la amplitud (0.038) se obtiene un resultado de 1.828 pero para fines práctico se puede redondear al valor de 1.83 y no hay mayor inconveniente.

Intervalo de Clase $(lim_{inf} - lim_{sup})$		Frecuencia Absoluta (f_i)	Frecuencia Absoluta Acumulada (F_i)	Frecuencia Relativa (f_R)	Frecuencia Relativa Acumulada (F_R)
1.60	1.638	3	3	0.075	0.075
1.638	1.676	6	9	0.15	0.225
1.676	1.714	2	11	0.05	0.275
1.714	1.752	8	19	0.20	0.475
1.752	1.79	9	28	0.225	0.7
1.79	1.83	12	40	0.30	1
n		40	$\sum f_R$	1.0	

Para calcular la frecuencia se cuentan los datos que caen dentro de la clase mayores que el límite inferior y nunca mayores o iguales que el límite superior, para la clase de 1.60 a 1.638 hay 3 datos 1.60, 1.62 y 1.62 pero por ejemplo si existiera 1.6381 ese ya no se toma en cuenta para esta clase, se considera para la siguiente. De igual forma se van contando todos los valores que entran en la descripción que dimos antes para cada clase. Se volverá a copiar la tabla de datos ordenados de forma ascendente y se resaltaran los números en color para determinar cuales corresponden a cada clase.

1.60	1.62	1.62	1.65	1.65	1.65	1.66	1.67	1.67	1.70
1.70	1.72	1.72	1.72	1.73	1.73	1.73	1.75	1.75	1.76
1.76	1.76	1.76	1.76	1.76	1.77	1.77	1.78	1.79	1.80
1.80	1.82	1.82	1.82	1.82	1.82	1.82	1.83	1.83	1.83

1.3 Medidas de tendencia central para **Datos Agrupados.**

Las medidas de tendencia central de la media, moda y mediana para datos agrupados, también llamados media, moda y mediana analítica, se obtienen de la siguiente manera:

1.3.1 Media analítica

Será igual a la sumatoria de los valores de la frecuencia por la marca de clase, entre el número de datos, esto se expresa matemáticamente así:

$$\bar{x} = \frac{\sum(f_i * X_m)}{n}$$

$\bar{x}$: media aritmética.
$\sum(f_i * X_m)$: se obtiene de la tabla de frecuencias.
n: número total de datos.

1.3.2 Moda

Para obtener la moda analítica o moda de valores agrupados se utiliza la siguiente ecuación:

$$m_0 = L_i + \left(\frac{(f_i - f_{i-1})}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \times A\right)$$

Dónde:

L_i: Límite inferior de la clase modal

$f_i - f_{i-1}$: es la diferencia de la frecuencia de la clase modal menos la anterior.

$f_i - f_{i+1}$: es la diferencia de la frecuencia de la clase modal menos la siguiente.

A: amplitud

La clase modal es aquella que tiene la frecuencia más alta. Estos valores se obtienen de la tabla de distribución de frecuencias, primero establecemos **cual es la clase modal.**

1. 3. 3 Mediana

La mediana para datos agrupados se calcula por medio de una expresión matemáticamente enunciada así:

$$m_e = L_i + \left(\frac{\frac{N}{2} - F_{(i-1)}}{f_i} * A \right)$$

Donde:

L_i: Límite inferior de la clase mediana

N: números de datos

$F_{(i-1)}$: Frecuencia acumulada anterior.

f_i: frecuencia de la clase mediana

A: amplitud

GRAFICOS.

Para una mejor interpretación de los datos se pueden crear gráficos. Los gráficos condensan una gran cantidad de información en un espacio reducido y esto facilita la lectura y asimilación de datos de forma rápida y simple.

Gráfico de Barras. Como su nombre lo dice están constituido por barras verticales que se trazan con la misma separación entre sí, dichas barras se construyen sobre un sistema de ejes, en el eje X se coloca la información de los datos o clases y en el eje Y la información de las frecuencias. Su principal función es comparar frecuencias de variables cuantitativas.

Pasos para construir un gráfico de barras:

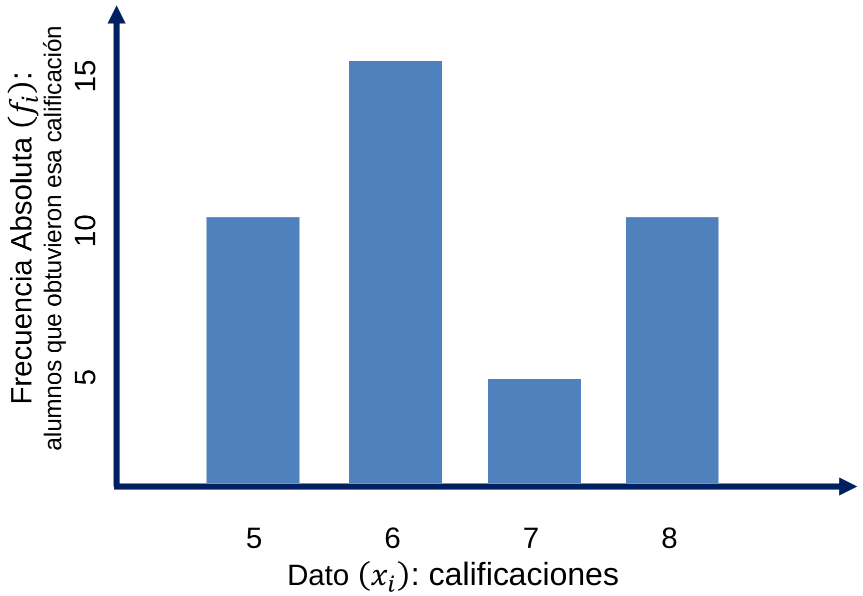
a) Se parte de un sistema de coordenadas rectangulares, en el eje X se coloca la variable de estudio y en el eje Y las frecuencias correspondientes.

b) Se construyen los rectángulos tomando como base el eje X y la altura de cada columna se determina con el valor de la frecuencia para cada valor.

c) El ancho de las barras es siempre el mismo mientras que el valor que varía es la altura de las mismas.

Vamos a tomar como ejemplo la tabla de frecuencias que construimos para datos no agrupados:

Dato (x_i)	Frecuencia Absoluta (f_i)	Frecuencia Absoluta Acumulada (F_i)	Frecuencia Relativa (f_R)	Frecuencia Relativa Acumulada (F_R)
5	10	10	0.25	0.25
6	15	25	0.375	0.625
7	5	30	0.125	0.75
8	10	40	0.25	1.0
N	40	$\sum f_R$	1.0	

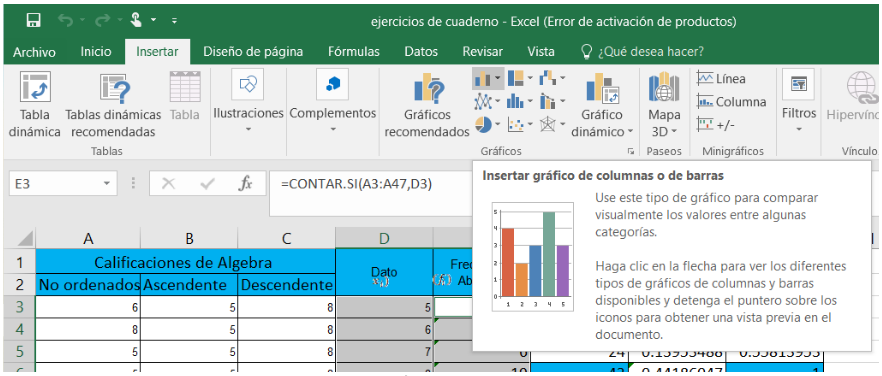


Si revisamos los pasos de construcción podemos ver que se cumplió con cada uno de ellos. Los datos en los ejes son correctos, existe la misma separación entre las barras y el ancho de cada una es igual.

En Excel se realiza de la siguiente forma:

a) Con el cursor seleccionan los datos numéricos correspondientes a Dato (x_i), se teclea CTRL y manteniendo la tecla presionada se seleccionan los datos de Frecuencia Absoluta (f_i).
b) Con ayuda del cursor nos desplazamos a la pestaña insertar y cuando se desplieguen los iconos buscamos el que nos aparezca la leyenda “Insertar Gráfico de Columnas o Barras.

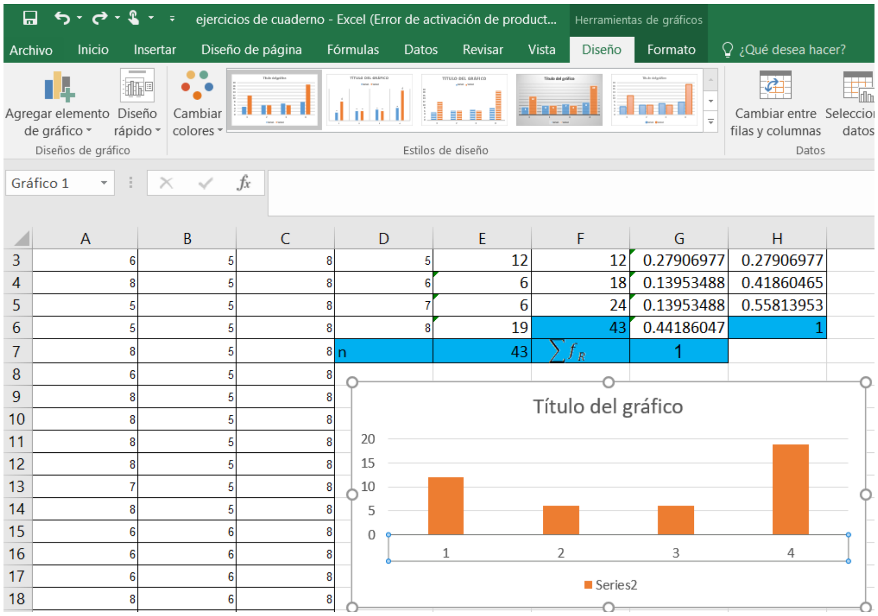
Imagen No. 19 Ejercicio



Fuente: Elaboración propia, noviembre 2020

c) Al dar clic sobre el ícono nos desplegar el gráfico en la hoja de trabajo, aparecerán 2 series debemos eliminar la que no corresponde a los datos de frecuencia, por lo general se elimina la serie 1. Para eliminarla se selecciona una de las columnas de la serie y se oprima la tecla SUPRIMIR. Y se obtiene lo siguiente:

Imagen No. 20 Ejercicio

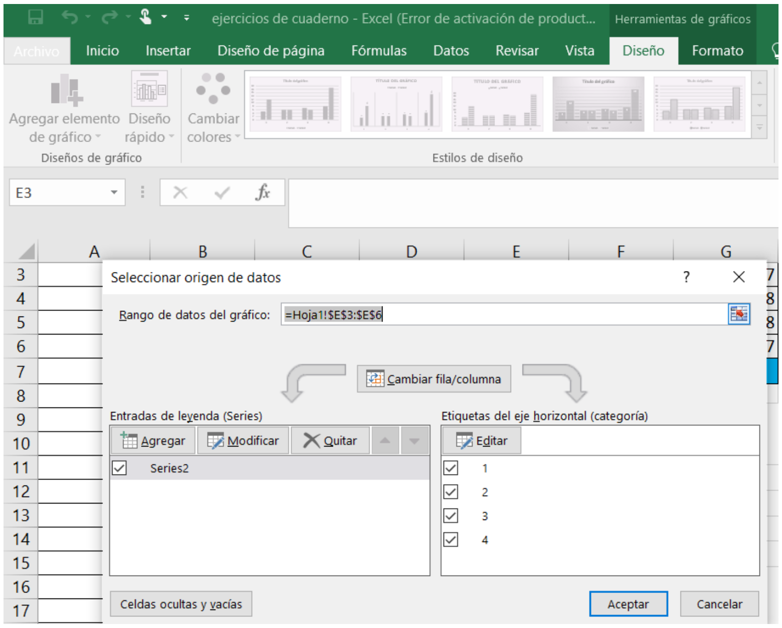


Fuente: Elaboración propia, noviembre 2020

d) En dato caso de que los datos del eje x no correspondan a los que están en nuestra tabla se pueden modificar dando clic derecho con el cursor sobre la palabra serie 2. En el submenú aparece “seleccionar datos” se da clic y aparece la siguiente subpantalla, en el recuadro izquierdo se puede modificar el nombre de la serie en lugar de “Series2” pondremos “datos” y en el recuadro derecho se pueden seleccionar los datos de las calificaciones.

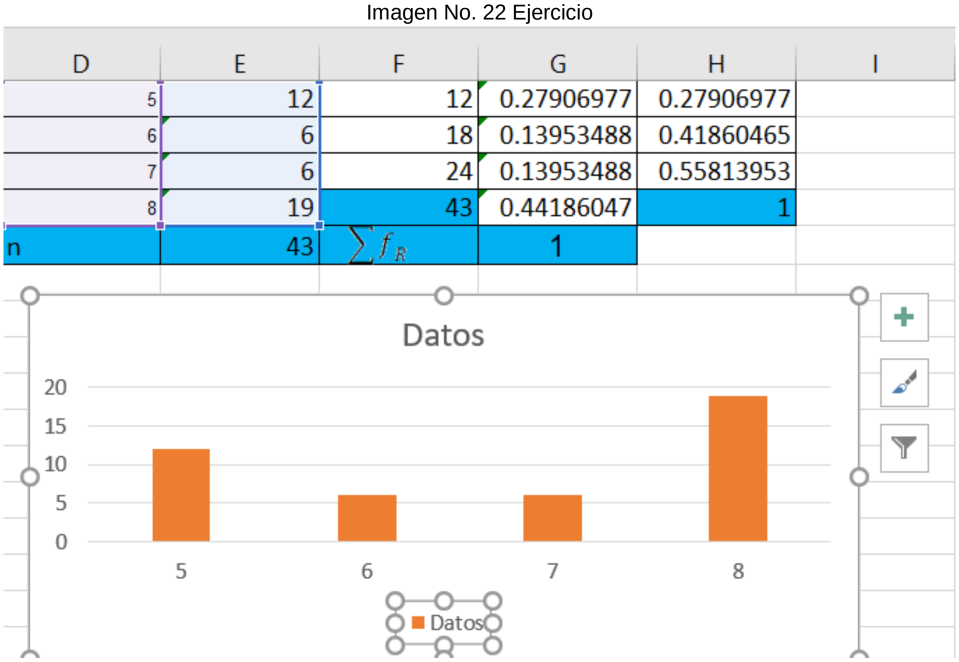
Cuando se haya realizado simplemente se da clic en aceptar y todo se reflejará en el gráfico.

Imagen No. 21 Ejercicio



Fuente: Elaboración propia, noviembre 2020

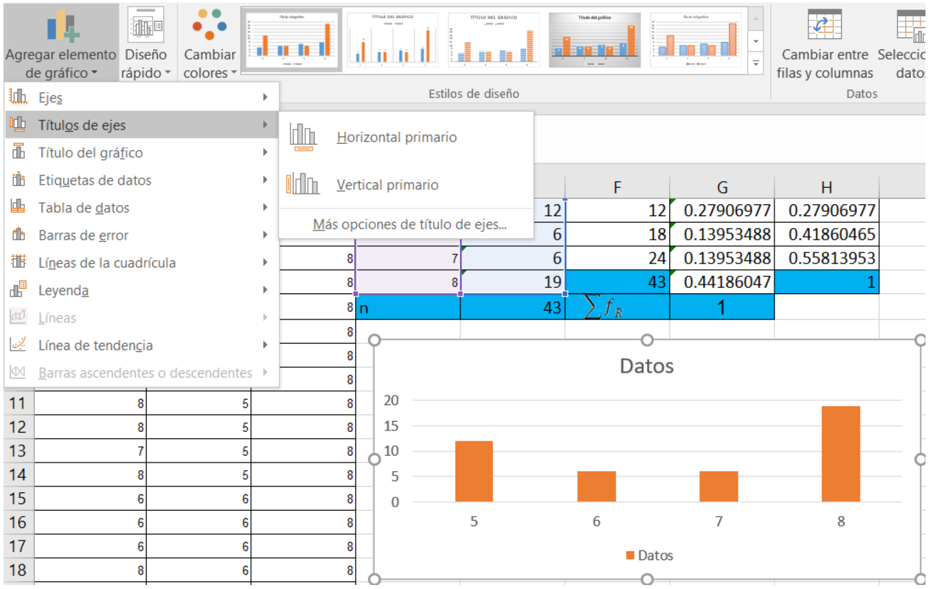
Imagen No. 22 Ejercicio



Fuente: Elaboración propia, noviembre 2020

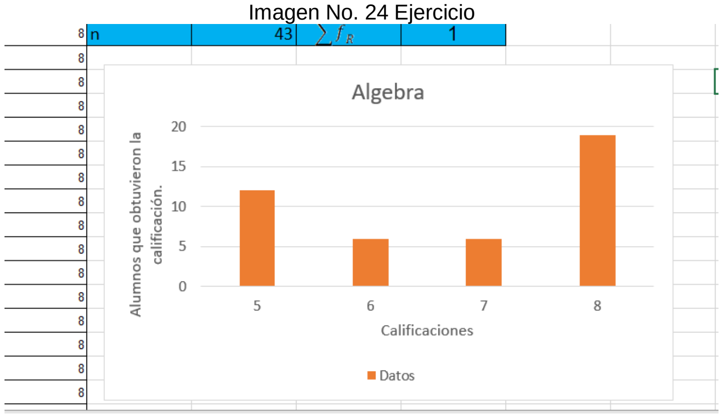
e) Por ultimo en el menú formato que se pueden modificar los datos generales de los ejes y el título del gráfico los cuales deben quedar como sigue: Eje x "Calificaciones" Eje y "Alumnos que obtuvieron la calificación" y en el Título del Grafico "Algebra".

Imagen No. 23 Ejercicio



Fuente: Elaboración propia, noviembre 2020

Imagen No. 24 Ejercicio



Fuente: Elaboración propia, noviembre 2020

Histograma. Es una representación gráfica para datos agrupados, es un conjunto de barras rectangulares verticales en el cual la altura de las mismas depende de los valores de la frecuencia absoluta. Los intervalos de clase se colocan en el eje x sin cortarse, a diferencia del gráfico de barras que si requiere distancia entre cada barra. La base de cada barra es proporcional a la amplitud de la clase.

Para construir el histograma se colocan en el eje “x” los intervalos de clase y en el eje “y” las frecuencias absolutas.

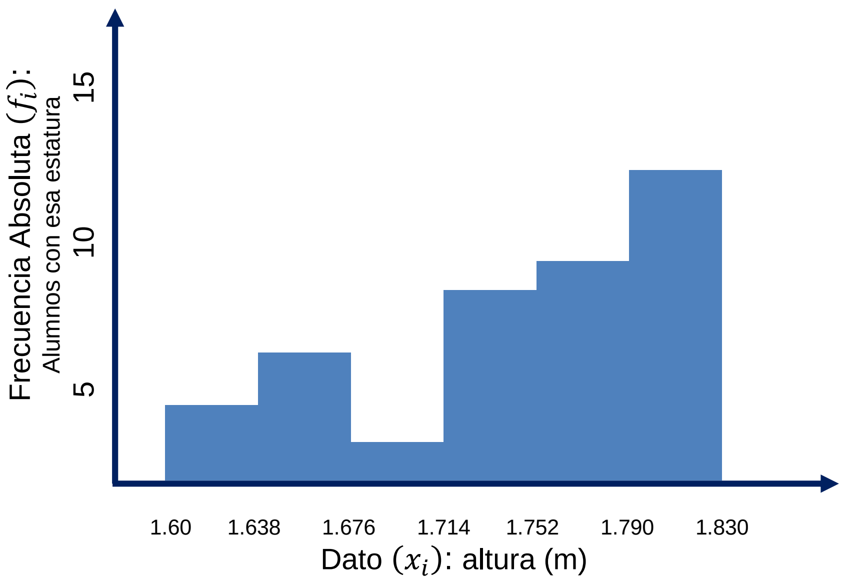
Se retomará el ejemplo de las alturas:

Intervalo de Clase $(lim_{inf} - lim_{sup})$		Frecuencia Absoluta (f_i)	Frecuencia Absoluta Acumulada (F_i)	Frecuencia Relativa (f_R)	Frecuencia Relativa Acumulada (F_R)
1.60	1.638	3	3	0.075	0.075
1.638	1.676	6	9	0.15	0.225
1.676	1.714	2	11	0.05	0.275
1.714	1.752	8	19	0.20	0.475
1.752	1.79	9	28	0.225	0.7
1.79	1.83	12	40	0.30	1
n		40	$\sum f_R$	1.0	

a) En el eje "x" se colocan los intervalos de clase empezando desde la clase con el dato menor hasta la clase con el dato mayor (también se puede utilizar la marca de clase)

b) En el eje "y" se coloca una marca para la frecuencia que corresponde al intervalo de clase.

c) Se dibujan las barras rectangulares de igual ancho con su respectiva altura que corresponde a la marca de la frecuencia absoluta. No debe haber separación entre las columnas.

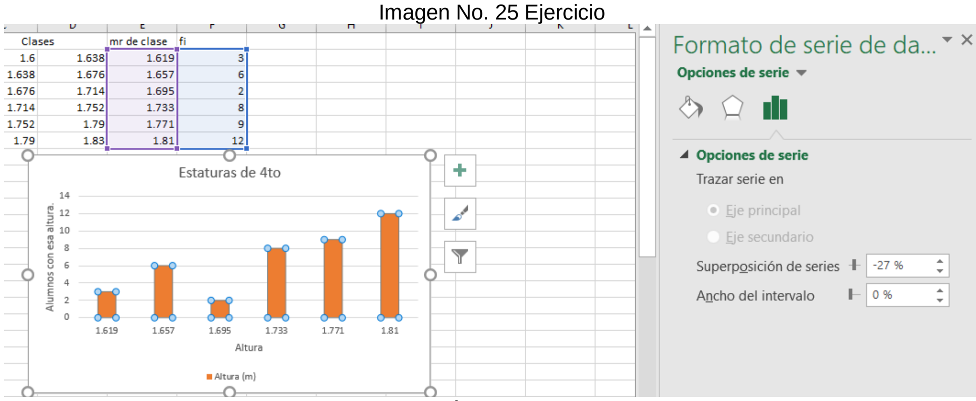


Si revisamos los pasos de construcción podemos ver que se cumplió con cada uno de ellos. Los datos en los ejes son correctos, no existe separación entre las barras y el ancho de cada una es igual.

En Excel se construye de la misma forma que la gráfica de barras, pero el final se hace lo siguiente.

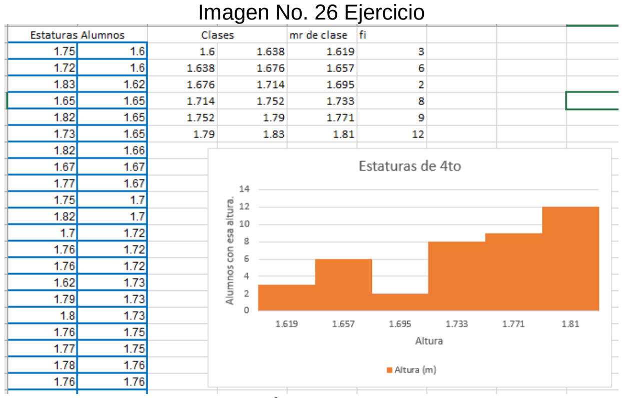
Se da clic sobre alguna columna y automáticamente se seleccionarán todas, dar clic en el botón derecho del mouse y en el menú que aparece dar clic sobre dar formato a serie de datos, se abrirá una ventana donde "superposición de series" y "ancho de intervalo" ambas con una barra de desplazamiento, en el ancho de intérvalo debe quedar en "0" y de esa forma ya no existe separación entre columnas.

Imagen No. 25 Ejercicio



Fuente: Elaboración propia, noviembre 2020

Imagen No. 26 Ejercicio



Fuente: Elaboración propia, noviembre 2020

Gráfico de Sectores Circulares

Es también llamado gráfico de pastel o gráfico circular sirve para representar variables cualitativas o discretas, representa la proporción de elementos de cada uno de los valores de la variable.

Se utilizan los valores de la Frecuencia Relativa que son proporcionales a los a cada parte del círculo.

Para construir la gráfica se debe partir de un círculo y calcular el ángulo de cada porción usando la siguiente fórmula:

$$\propto = 360° * f_r$$

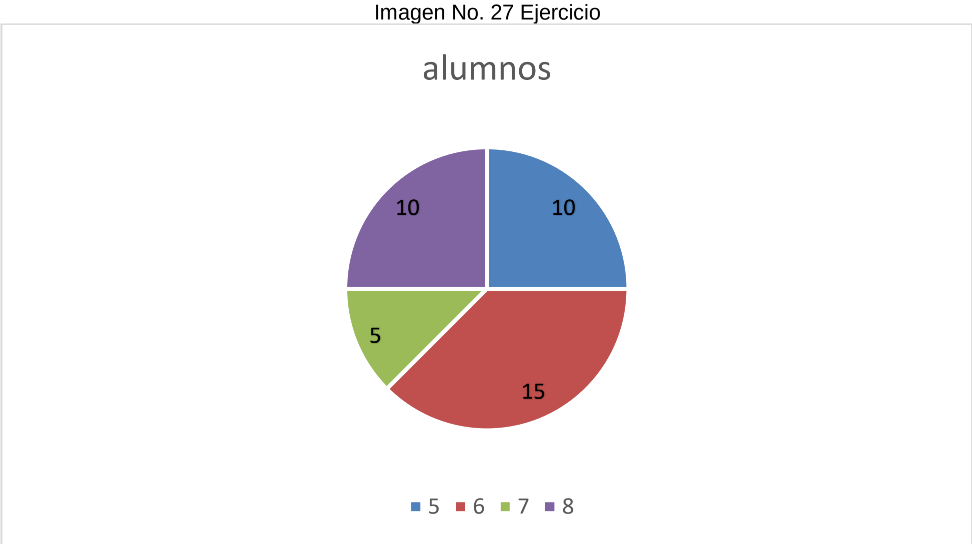
Una vez que ya se tienen los valores de los ángulos de dibuja un círculo con la ayuda de un compás o algún objeto circular y para medir los ángulos se usa el transportador. Se parte desde 0° y se sigue el sentido de las manecillas del reloj.

Para expresar los datos de cada porción se multiplica cada valor de la frecuencia relativa por 100.

Vamos a retomar el ejemplo de las calificaciones:

Dato (x_i)	Frecuencia Absoluta (f_i)	Frecuencia Absoluta Acumulada (F_i)	Frecuencia Relativa (f_R)	Ángulo de Cada Porción $(\propto = 360° * f_R)$
5	10	10	0.25	90
6	15	25	0.375	135
7	5	30	0.125	45
8	10	40	0.25	90

Imagen No. 27 Ejercicio

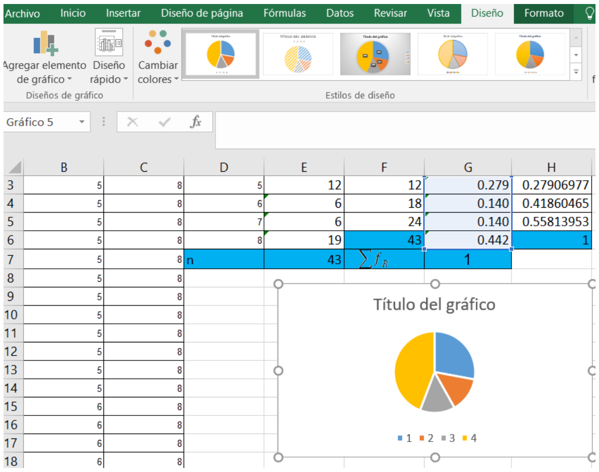


Fuente: Elaboración propia, noviembre 2020

En Excel o en un programa de hojas de cálculo:

1. Se selecciona la columna frecuencia relativa.
2. En la pestaña de insertar se ubica el icono de insertar gráfico circular o de anillos y seleccionar el primer icono de gráficos 2D.

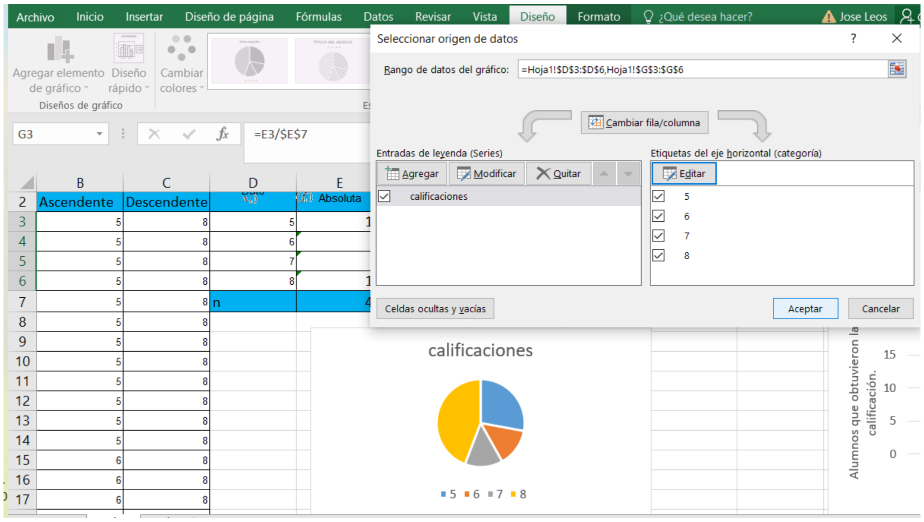
Imagen No. 28 Ejercicio



Fuente: Elaboración propia, noviembre 2020

3. Dar clic con el botón derecho del mouse teniendo el gráfico seleccionado y dar clic en "seleccionar datos, en el recuadro donde dice "entrada de leyendas" seleccionar los valores de la columna de la "frecuencia relativa" y en "etiquetas de eje" seleccionar los valores de la columna "dato", dar en aceptar y se desplegar el gráfico.

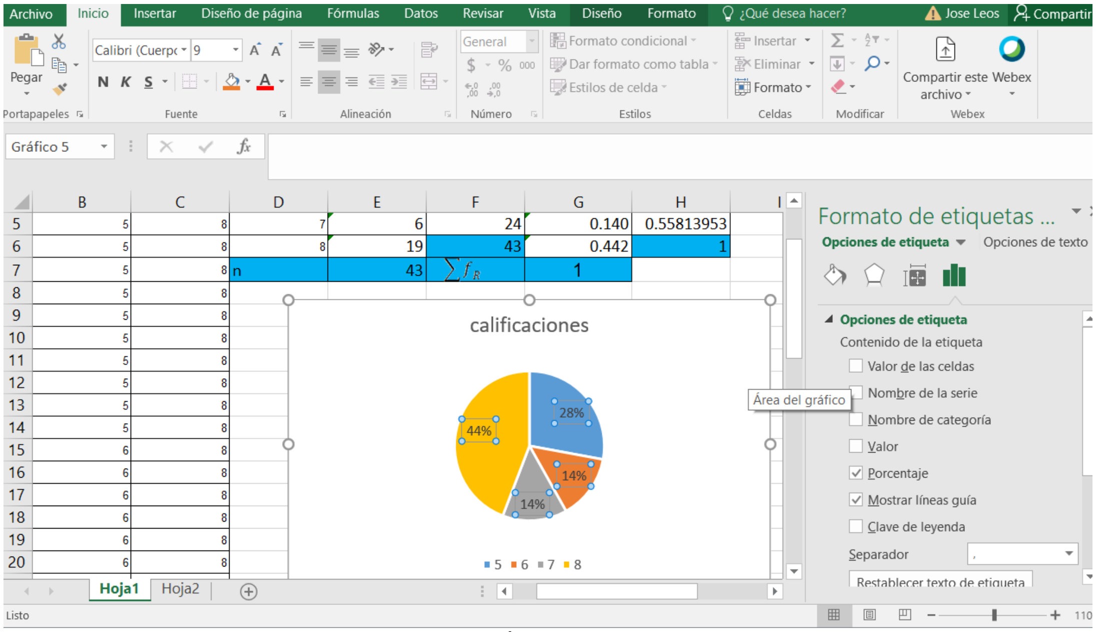
Imagen No. 29 Ejercicio



Fuente: Elaboración propia, noviembre 2020

4. Dar clic con el botón derecho del mouse y seleccionar "Agregar Etiquetas de Datos" y luego de nuevo usando el botón derecho del mouse seleccionar "Formato de Etiqueta de Datos", dejar de marcar la casilla de verificación de "valor" y marcar "porcentaje" y la gráfica queda lista con los datos mostrados sobre el gráfico en porcentajes.

Imagen No. 30 Ejercicio

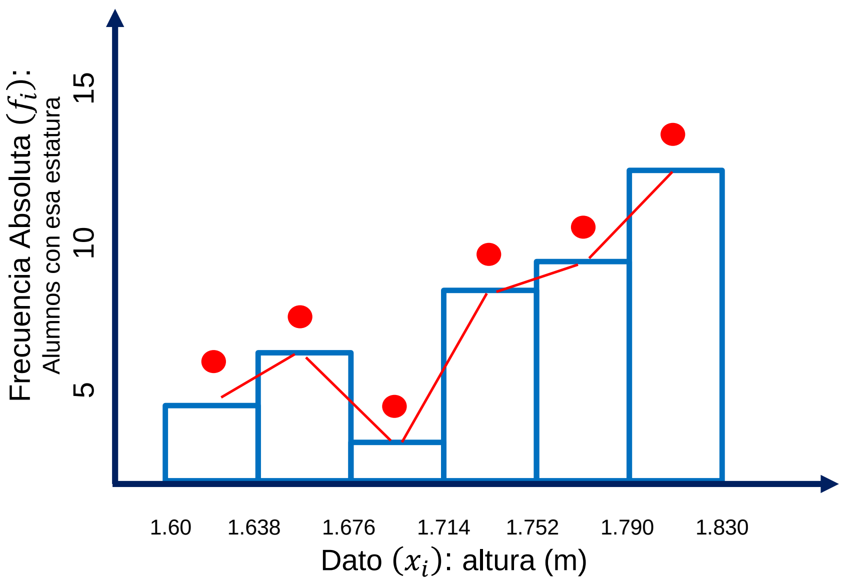


Fuente: Elaboración propia, noviembre 2020

Polígono de Frecuencias.

Es un tipo de gráfico que surge a partir de un polígono de frecuencias, para ello se localizan los puntos medios de cada columna presente el histograma y se unen con líneas rectas formando una figura geométrica o polígono.

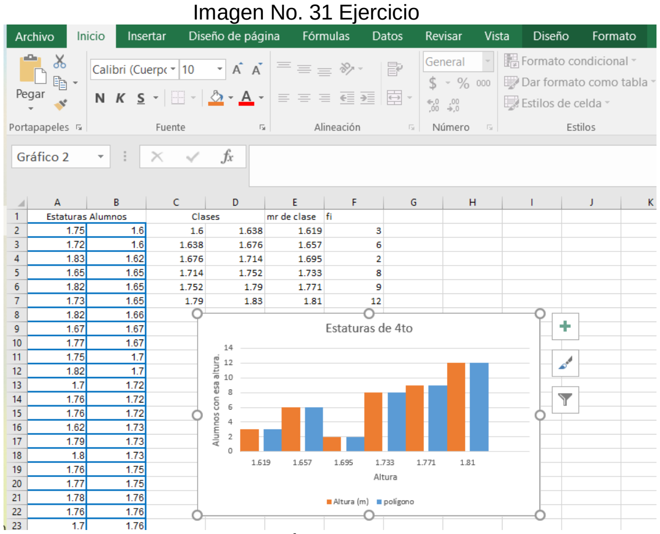
Vamos a retomar el histograma que ya se ha realizado previamente.



En Excel o en un programa que genere hojas de cálculo se hace como sigue:

1. Construir el Histograma (en este caso se tomará el del ejemplo resuelto)
2. Agregar una serie al gráfico, para ello damos clic en el botón derecho del mouse y seleccionamos “seleccionar origen de datos” en el recuadro “entrada de leyenda” se selecciona agregar y se le pone el nombre de “polígono” y se selecciona la columna de valores de la frecuencia absoluta y se le da clic en aceptar. Se duplicará el mismo gráfico.

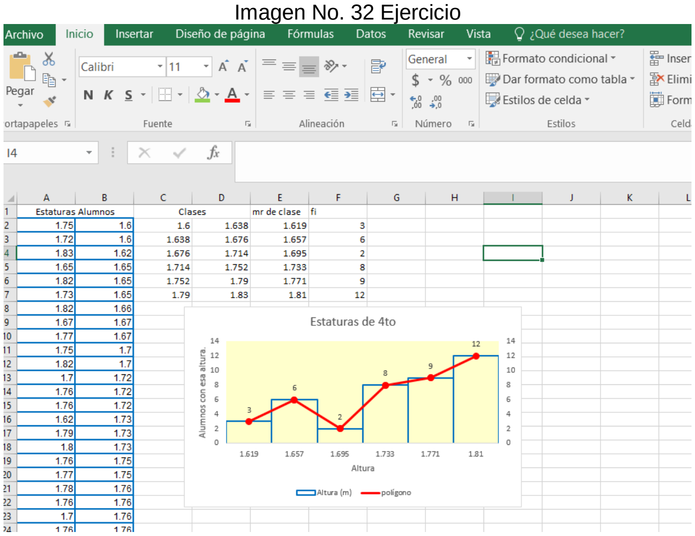
Imagen No. 31 Ejercicio



Fuente: Elaboración propia, noviembre 2020

3. Con el gráfico seleccionado dar clic derecho al mouse y seleccionar “cambiar tipo de gráfico” en la ventana que se genera seleccionar la opción “cuadro combinado” y luego el segundo icono que aparece. En la sección del formato se elimina el relleno para las barras del histograma y se deja sólo el contorno.

Imagen No. 32 Ejercicio



Fuente: Elaboración propia, noviembre 2020

Ejercitando mi habilidad. (Productos esperados)

Realiza los siguientes ejercicios.

1. La pediatra obtuvo la siguiente tabla sobre los meses de edad de 50 niños de su consulta en el momento de andar por primera vez.

Meses	Niños
9	1
10	4
11	9
12	16
13	11
14	8
15	1

a) Construir la tabla de frecuencias usando como referencia la tabla en blanco.

Dato (x_i)	Frecuencia Absoluta (f_i)	Frecuencia Absoluta Acumulada (F_i)	Frecuencia Relativa (f_R)	Frecuencia Relativa Acumulada (F_R)
N		$\sum f_R$		

b) Realiza una Gráfica de Barras.

c) Calcular la moda, la mediana, la media y la varianza.

2. A un grupo de alumnos del 3° Semestre se les aplico una prueba de condición física y se obtuvieron los siguientes resultados que son las velocidades promedio:

12	6	2	5	3	16	13
18	12	1	10	21	11	19
17	14	19	18	4	16	15
14	4	2	15	20	10	20
5	2	15	4	11	7	9
1	9	9	9	5	1	17
6	21	9	21	17	18	9
5	16	19	3	21	20	15
17	5	12	18	8	13	19

a) Ordenar los datos en forma ascendente.

b) Calcula el número de clases aplicando la regla de Sturges

K = 1 + 3.3 log(N) =

c) Determina el Rango

R = $X_{\text{Dato mayor}} - X_{\text{Dato menor}} =$

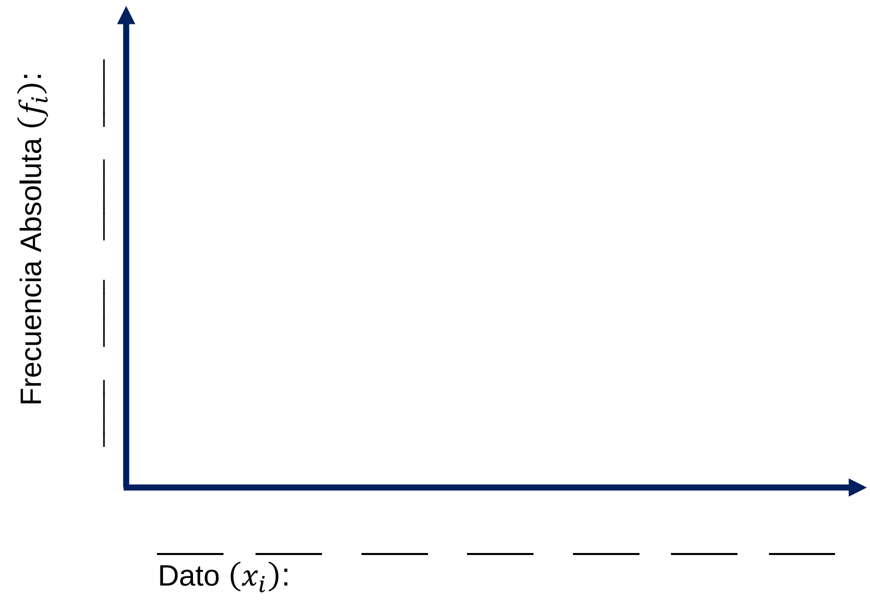
d) Calcula la Amplitud de la clase

A = $\frac{R}{K} =$

e) Construye la tabla de frecuencias usando como referencia la tabla en blanco.

Intervalo de Clase $(lim_{inf} - lim_{sup})$		Frecuencia Absoluta (f_i)	Frecuencia Absoluta Acumulada (F_i)	Frecuencia Relativa (f_R)	Frecuencia Relativa Acumulada (F_R)
N			$\sum f_R$		

f) Realiza un Histograma y el polígono de frecuencias.



g) Calcular la moda, la mediana, la media y la varianza.

¿Qué Aprendí?

1. Elabora la tabla de distribución de Frecuencias.
2. Realiza una representación gráfica (HISTOGRAMA, POLÍGONO DE FRECUENCIAS Y SECTOR CIRCULAR).
3. Calcula las medidas de tendencia central para Datos Agrupados

- Media
- Moda
- Mediana

CLASE	L_i	L_s	X_m	f_i	F_i	n_i	N_i	$N_i = \%$	$X_m * f_i$
1	[31.5	42.5)		7					
2	[42.5	53.5)		18					
3	[53.5	64.5)		16					
4	[64.5	75.5)		6					
5	[75.5	86.5)		5					
6	[86.5	97.5)		2					
7	[97.5	108.5)		1					
				∑=55		∑=			∑=

Las velocidades, en kilómetros por hora de 30 automóviles rastreados de manera aleatoria en una avenida de la ciudad de Celaya, Gto. Aparecen a continuación:

Clase	Li	Ls	Xm (MARCA CLASE)	Fi	Frec Acu. **(Fi)**	Frec. Rela (**ni**)	Frec. Rela. Acumu(**Ni**)

66	75	61	72	74
78	64	77	51	58
80	76	67	78	74
79	70	57	72	79
71	62	73	59	67
66	68	60	68	65

1. Completa la distribución de frecuencias.

2. ¿Cuál es el promedio de la velocidad de los automóviles? ¿Qué medida de tendencia central tuviste que calcular para saber este dato? ¿Por qué?

3. Si el límite de velocidad es de 70 km/hrs. ¿Cuántos automóviles viajan a la velocidad adecuada? ¿Por qué?

4. Si existiera operativo por parte de la Policía Vial ¿Cuántos vehículos serían infraccionados por no respetar el límite de velocidad?

Rescatando mis Aprendizajes

Se le pide amablemente consulte los siguientes sitios web donde encontrarás videos con lo que ya se abordó en la unidad y que te servirán como referencia futura.

https://es.khanacademy.org/math/statistics-probability/summarizing-quantitative-data/mean-median-basics/v/statistics-intro-mean-median-and-mode

https://es.khanacademy.org/math/cc-sixth-grade-math/cc-6th-data-statistics/dot-plot/v/ways-to-represent-data

https://es.khanacademy.org/math/cc-sixth-grade-math/cc-6th-data-statistics/dot-plot/v/ways-to-represent-data

UNIDAD III

UNIDAD III.

La Estadística proporciona una serie de principios, procedimientos, técnicas y métodos para cuatro tareas fundamentales en la investigación social y los estudios técnicos: • Obtener datos pertinentes de manera rápida y a costos bajos. • Los métodos para su organización y procesamiento, a fin de obtener de ellos la información requerida. • Los principios y métodos para que las conclusiones emanadas o acciones a seguir sean producto de procesos de inducción válidos, que se obtengan de interpretaciones adecuadas de los resultados.

Una de las medidas de posición y la forma de encontrarlos en datos no agrupados, dentro del programa de estudio es el de cuartiles, es y percentiles. En esta unidad también verás temas de:

- Medidas de dispersión
- Cuartiles es y percentiles
- Regresión lineal

¡Ya vamos muy avanzados en la asignatura, anímate!!! Esta es la última parte donde aprenderás situaciones de la vida real en lo cotidiano y en lo laboral.

Para aprender más

MEDIDAS DE DISPERSIÓN

Parámetros estadísticos que permiten conocer la dispersión de los datos, es decir cómo se alejan los datos respecto de la media aritmética. Sirven como indicador de la variabilidad de los datos. Las medidas de dispersión más utilizadas son el rango, la desviación estándar y la varianza.

Rango

Indica la dispersión entre los valores extremos de una variable. se calcula como la diferencia entre el mayor y el menor valor de la variable. Se denota como R.

Para datos ordenados se calcula como:

$$R = x_n - x_1$$

Donde:

x_n: Es el mayor valor de la variable.

x_1: Es el menor valor de la variable.

Media

Es la media aritmética de los valores absolutos de las diferencias de cada dato respecto a la media.

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

Donde:

x_i: valores de la variable.

n: número total de datos

Desviación estándar

La desviación estándar mide el grado de dispersión de los datos con respecto a la media, se denota como s para una muestra o como σ para la población. Se define como la raíz cuadrada de la varianza según la expresión:

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

Obsérvese que el denominador es n - 1, a diferencia de la desviación media donde se divide entre n; también existe la fórmula de desviación típica donde el denominador es n pero se prefiere n-1.

Mientras menor sea la desviación estándar, los datos son más homogéneos, es decir existe menor dispersión, el incremento de los valores de la desviación estándar indica una mayor variabilidad de los datos.

Varianza

Es otro parámetro utilizado para medir la dispersión de los valores de una variable respecto a la media. Corresponde a la media aritmética de los cuadrados de las desviaciones respecto a la media. Su expresión matemática es:

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

Donde: x_i es el dato i-ésimo y $\bar{x}$ es la media de los n datos

Los cálculos de las estadísticas pueden resultar sencillos si el conjunto de datos es pequeño

Un conjunto de datos ya de cierto tamaño hace que los cálculos resulten lentos y con probabilidades de errores al operar con ellos.

Del mismo modo que para la media, no siempre será posible encontrar la varianza, y es un parámetro muy sensible a las puntuaciones extremas. Se puede observar que, al estar la desviación elevada al cuadrado, la varianza no puede tener las mismas unidades que los datos.

Comparando con el mismo tipo de datos, una varianza elevada significa que los datos están más dispersos. Mientras que un valor de la varianza bajo indica que los valores están por lo general más próximos a la media.
Un valor de la varianza igual a cero implica que todos los valores son iguales, y por lo tanto también coinciden con la media aritmética.

Ejemplo
En un partido de baloncesto, se tiene la siguiente anotación en los jugadores de un equipo: 0,2,4,5,8,10,10,15,38. Calcular la varianza de las puntuaciones de los jugadores del equipo.

Aplicando la fórmula

$$\bar{x} = \frac{0+2+4+5+8+10+10+15+38}{9} = \frac{92}{9} = 10.22$$

Se obtiene la media.

Seguidamente se aplica la fórmula de la varianza:

$$\sigma = \frac{\begin{array}{c}(0-10.22)^2+(2-10.22)^2+(4-10.22)^2+(5-10.22)^2+(8-10.22)^2+\\(10-10.22)^2+(10-10.22)^2+(15-10.22)^2+(38-10.22)^2\end{array}}{9}$$

$$\sigma = \frac{\begin{array}{c}(-10.22)^2+(-8.22)^2+(-6.22)^2+(-5.22)^2+(-2.22)^2+\\(-0.22)^2+(-0.22)^2+(4.78)^2+(27.78)^2\end{array}}{9}$$

$$\sigma = \frac{104.4484+67.5684+38.6884+27.2484+4.9284+0.0484+22.8484+771.7284}{9}$$

$$\sigma = \frac{1037.55569}{9} = 115.28$$

$$\sigma = 115.28$$

Cálculo de la varianza para datos agrupados

En el caso de N muestras agrupadas en n clases se aplica la fórmula que queda simplificada como:

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

La interpretación que se puede hacer del resultado es la misma que para datos no agrupados.

Ejemplo

La altura en cm de los jugadores de un equipo de baloncesto está en la siguiente tabla. Calcular la varianza.

Intervalos de Clase	*Xi*	*Fi*
[160,170)	165	1
[170,180)	175	2
[180,190)	185	4
[190,200)	195	3
[200,210)	205	2

En primer lugar, rellenar la siguiente tabla:

Intervalos de clase		$\bar{x}_i$	fi	$\bar{x}_i f_i$		$x_i - \mu$	$(x_i - \mu)^2$	$f_i(x_i - \mu)^2$
160	170	165	1	165		-22.5	506.25	506.25
170	180	175	2	350		-12.5	156.25	312.5
180	190	185	4	740		-2.5	6.25	25
190	200	195	3	585		7.5	52.65	168.75
200	210	205	2	410		17.5	306.25	612.5
μ		n	12	$\sum f_i\bar{x}_i$	2250		$\sum (f_i\bar{x}_i)^2$	1625

Se debe calcular la media para poder aplicar la fórmula.

$$\bar{\mu} = \frac{\sum f_i \bar{x}_i}{N} = \frac{2250}{12}$$
$$\bar{\mu} = 187.50$$

Se calcula entonces la varianza y la desviación estándar.

$$\sigma^2 = \frac{\sum_{i=1}^{n} f_i (x_i - \mu)^2}{N}$$
$$\sigma^2 = \frac{1625}{12}$$
$$\sigma^2 = 135.41$$

Propiedades de la varianza

1. σ^2 ≥ La varianza es un valor positivo, como ya se ha comentado anteriormente, la igualdad sólo se da en el caso de que todas las muestras sean iguales.
2. Si a todos los datos se les suma una constante *(k),* la varianza sigue siendo la misma.
3. Si todos los datos se multiplican por una constante, la varianza queda multiplicada por el cuadrado de la constante.
4. Si se disponen de varias distribuciones con la misma media y se calculan las distintas varianzas, se puede hallar la varianza total aplicando la fórmula:

$$\sigma2=\sigma21+\sigma22+\ldots+\sigma2nn$$

En el caso de que las distribuciones tengan distinto tamaño, la fórmula se pondera y queda como:

$$\sigma2=\sigma21k1+\sigma22k2+\ldots+\sigma2nknk1+k2+\ldots+kn$$

Ejemplo

En un examen, todos los alumnos de la clase sacaron un diez. Hallar la varianza de las notas.

Al coincidir todos los valores la media coincide también con ellos $\bar{x}=10$, y la varianza es nula $\sigma^2=0$.

Desviación típica

La desviación típica es la raíz cuadrada de la varianza y se representa por la letra σ. Para calculara se calcula la varianza y se saca la raíz. Las interpretaciones que se deducen de la desviación típica son, por lo tanto, parecidas a las que se deducían de la varianza.

Comparando con el mismo tipo de datos, una desviación típica elevada significa que los datos están dispersos, mientras que un valor bajo indica que los valores son próximos los unos de los otros, y por lo tanto de la media.

Propiedades de la desviación típica

1. $\sigma\geq 0$ La desviación típica es un valor positivo, la igualdad sólo se da en el caso de que todas las muestras sean iguales.
2. Si a todos los datos se les suma una constante, la desviación típica sigue siendo la misma.
3. Si todos los datos se multiplican por una constante, la desviación típica queda multiplicada por dicha constante.
4. Si se dispone de varias distribuciones con la misma media y se calculan las distintas desviaciones típicas, se puede hallar la desviación típica total aplicando la fórmula:

$$\sigma=\sigma21+\sigma22+\ldots+\sigma2nn----------------\surd$$

En el caso de que las distribuciones tengan distinto tamaño, la fórmula se pondera y queda como

$$\sigma=\sqrt{\frac{\sigma_1^2 k_1+\sigma_2^2 k_2+\ldots+\sigma_n^2 k_n}{k_1+k_2+\ldots+k_n}}$$

VARIANZA Y DESVIACIÓN ESTANDAR

Desviación Estándar (σ) mide cuanto se separan los datos.

La fórmula es fácil: es la raíz cuadrada de la varianza. Así que ¿Qué es la varianza?

Varianza es el cuadrado de la desviación estándar (σ^2) se define como:

La media de las diferencias con la media elevada al cuadrado

En otras palabras, sigue estos pasos:

1.- Calcula la media (el promedio de los números)

2.- Ahora, por cada número resta la media t eleva el resultado al cuadrado (la diferencia elevada al cuadrado).

3.- Ahora calcula la media de esas diferencias al cuadrado (¿Por qué al cuadrado?)

Ejemplo

Tú y tus amigos habéis medido las alturas de vuestros perros (en milimetros):

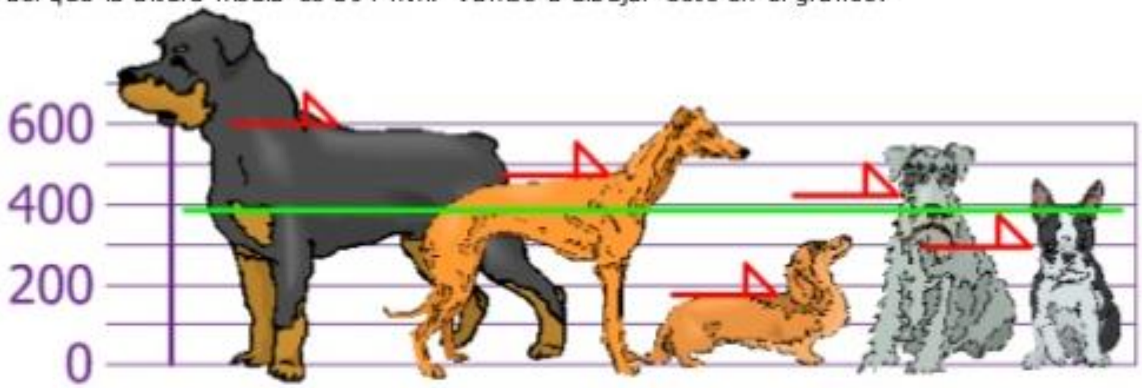
Las alturas (de los hombros) son: 600mm, 470mm, 170mm, 430mm y 300mm.

Calcula la media, la varianza y la desviación estándar.

Respuesta:

$$\text{Media} = \frac{600 + 470 + 170 + 430 + 300}{5} = \frac{1970}{5} = 394$$

así que la altura media es 394 mm. Vamos a dibujar esto en el gráfico:



Ahora calculamos la diferencia de cada altura con la media:

Para calcular la varianza, toma cada diferencia, elévala al cuadrado, y haz la media:

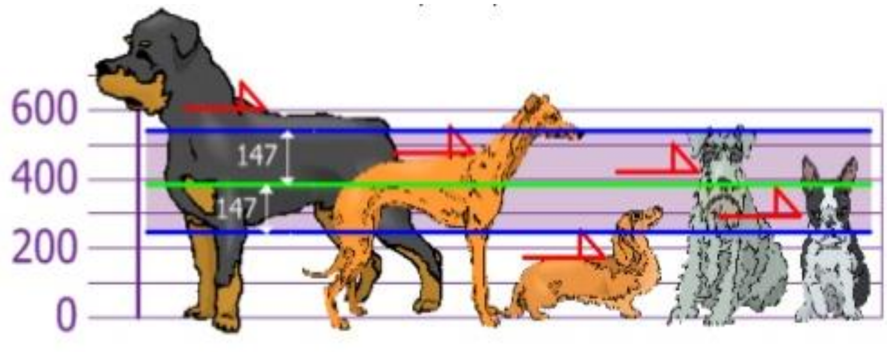
$$\text{Varianza: } \sigma^2 = \frac{206^2 + 76^2 + (-224)^2 + 36^2 + (-94)^2}{5} = \frac{108{,}520}{5} = 21{,}704$$

Así que la varianza es 21,704.

Y la desviación estándar es la raíz de la varianza, así que:

Desviación estándar: $\boldsymbol{\sigma = \sqrt{21{,}704} = 147}$

Y lo bueno de la desviación estándar es que es útil, ahora veremos que alturas están a distancia menos de la desviación estándar (147 mm) de la media:



Así que usando la desviación estándar tenemos una manera “estándar” de saber que es normal, o extra grande o extra pequeño.

Los Rottweiler son perros grandes. Y los Dachsund son un poco menudos… ¡pero que no se enteren!

***Nota: ¿por qué al cuadrado?**

Elevar cada diferencia al cuadrado hace que todos los números sean positivos (para evitar que los números negativos reduzcan la varianza) Y también hacen que las diferencias grandes destaquen. Por ejemplo $100^2 = 10{,}000$ es mucho más grande que $50^2 = 2{,}500$

Para elevarlas al cuadrado hace que la respuesta sea muy grande, así que lo deshacemos (con la raíz cuadrada) y así la desviación estándar es mucho más útil.

REGLAS DE SUMATORIAS

1.- SUMATORIA DE LOS DATOS DE UNA VARIABLE

Para hallar la sumatoria de los datos de una variable no hay más procedimiento que el de una agregaciones decir agregar a cada uno de los datos que sigue, hasta terminar, simbólicamente esta regla se escribe así $\sum x = X1 + X2 + X3 + \cdots XN$

2.- SUMATORIA DE UNA COSNTANTE

La sumatoria de una constante que aparece n veces en un conjunto es simplemente n veces constante $\sum C = C + C + C + \cdots + C$

$$\sum C = NC$$

sí TENEMOS 5 DE VALOR DE c Y SI c=410 entonces:

$$\sum C = NC * 5(10)$$

$$\sum C = 50$$

La diferencia entre la regla y la segunda es que la primera x es una variable que adopta diversos valores, mientras que la segunda c es 1 valor que no cambia.

3.- SUMATORIA DE UNA VARIABLE Y UNA CONSTANTE SUMADA O RESTADA.
Es igual a los datos de una variable más n veces la constante de la expresión
$\sum(X + c)$

4.- SUMATORIA DE UNA VARIABLE CON UN MULTIPLICADOR O UN DIVISOR
En su función de multiplicador o divisor, por la sumatoria de los datos de la variable
∑CX

5.- SUMATORIA DE POTENCIAS Y RAIZES DE UNA VARIABLE

Primero se halla la potencia o la raíz y luego se lleva a cabo la sumatoria, recordemos que con símbolo como. Es el cuadrado o segunda potencia de un valor x entonces, la expresión sumatoria de ∑ = indica elevar al cuadrado a cada dato de la variable X y posteriormente sumar las potencias, sustituyendo valores numéricos.

6.- REGLA PARA SUSTITUIR LA SUMATORIA

Si después de la sumatoria se encuentran entre paréntesis una expresión que incluye solo porciones de suma o resta la sumatoria puede estar distribuida entre los términos de la expresión ∑(X+Y.C= ∑X+∑Y.∑C, ∑(X+Y.C=∑X+∑Y-NC

Note que esta regla es válida para la variables o constantes.

7.- SUMATORIA DEL PRODUCTO O EL CONTANTE DE 2 O MAS VARIABLES

Esta expresión manda multiplica cada uno de los datos de la variable X por el que le corresponde en l variable Y y finalmente, sumar los productos.

Ejemplo:

datos:
x=8,12,16,20,24,28
28+3)
Y=4,6,8,10,12,14
Z=1,2,3,4,5,6
W=2,4,6,8,10,12
C=4, B=3

∑(X+B)= (8+39+(12+3)+(16+3)+(20+3)+(24+3)+(

=11+15+19+23+27+31
∑(X+B)= 126

∑yz=(4)(1)(6)(2)(8)(3)(10)(4)(12)(5)14)(6)

=4+12+24+400+60+84

∑yz=624

http://www.disfrutalasmatematicas.com/datos/desviacion-estandar.html

http://probabilidadyestadistica5acbtis.blogspot.mx/2010/12/reglas-de-la-sumatoria.html

VER EJERCICIOS DE: VARIANZA Y DESVIACIÓN TÍPICA

Ejercitando mi habilidad. (Productos esperados)

EJERCICIOS

Instrucciones. - Resuelve los siguientes ejercicios relativos a las medidas de dispersión, que te servirán como preparación para el examen parcial.

Envía los ejercicios resueltos al *portafolio.*

Ejercicio 1.

Dada la muestra 8, 3, 12, 7, 10, 6, 9 y 9: encuentra la media, la mediana, el rango, la varianza, la desviación estándar y el coeficiente de variación.

Ejercicio 2.

Dada la siguiente muestra de 20 medidas

8	**9**	**10**	**7**	**5**	**3**	**10**	**11**	**4**	**9**
7	6	0	12	8	16	4	-2	10	7

a) Construye un diagrama de puntos

b) Encuentra la media, la mediana, la varianza y la desviación estándar.

Ejercicio 3.

Una muestra de los recibos de ventas semanales para un bar está en cientos de dólares y se presenta a continuación:

43.3	**54.2**	**34.8**	**42.9**	**49.2**	**29.5**	**28.6**

Al ver estos resultados, los dueños implementan un programa publicitario diseñado para “emparejar” las ventas. Después de haber puesto en marcha el programa publicitario, se vuelve a tomar otra muestra y los resultados son los siguientes:

45.5	39.5	35.7	36.7	42.6	42.14

¿La campaña publicitaria logró su objetivo?

Ejercicio 4.

Adolfo Sánchez compró 20 acciones a $15 cada una; 50 acciones a $20 cada una; 100 acciones a $30 cada una y 75 acciones a $35 cada una.

a) ¿Cuál es el monto total de su inversión?

b) ¿Cuál es el precio promedio por acción?

Ejercicio 5.

Anabel Miranda utiliza dos máquinas diferentes para producir papeleras para las fotocopiadoras Kodak. Unas muestras de las papeleras de la primera máquina midieron 12.2, 11.9, 11.8, 12.1, 11.9, 12.4, 11.3 y 12.3 pulgadas. Las bandejas elaboradas con la segunda máquina midieron 12.2, 11.9, 11.5, 12.1, 12.2, 11.9, y 11.8 pulgadas. Anabel debe utilizar la máquina con la mayor consistencia en los tamaños de las papeleras. ¿Cuál máquina debe utilizar?

Ejercicio 6.

Los siguientes datos de muestras se han obtenido para el número de clientes diarios de la florería “El Tulipán negro”: 34, 45, 23, 34, 26, 32, 31, 41. Calcula la varianza, la desviación estándar y el rango.

Ejercicio 7.

La siguiente es una muestra de las ganancias por acción en dólares, para las acciones cotizadas en la bolsa de valores de Nueva York:

1.12	1.43	2.17	-1.19	2.87	-1.49

Considera que una cantidad negativa no es una ganancia, sino una pérdida. Calcula la varianza, la desviación estándar y el rango.

Ejercicio 8.

Las horas trabajadas por un empleado cada semana durante los dos últimos meses son:

52	48	37	54	48	15	42	12

Asumiendo que estos son datos muestrales, calcula

a) La media

b) La mediana

c) La moda

d) Para este caso ¿cuál medida consideras que representa mejor el punto central?

Ejercicio 9.

Utilizando las horas trabajadas por el empleado en el ejercicio anterior, calcula e interpreta.

a) El rango

b) La varianza

c) La desviación estándar

d) El primer cuartil

e) El percentil 25

Ejercicio 10.

Dados los siguientes puntajes de 9 pruebas para la clase de estadística, calcula el coeficiente de variación. Asume que son datos muestrales.

80	83	87	85	90	86	84	82	88

Ejercicio 11.

Los sindicalistas de la Ford Motor Company argumentan que, en contravención del contrato laboral, los trabajadores de la línea de producción hacen un promedio salarial tienen un promedio salarial menor (por hora) con una mayor variabilidad que los trabajadores de oficina. Una muestra de 10 elementos se toma de cada clase de trabajadores, arrojando los siguientes valores ¿Tales valores apoyan a los sindicalistas?

Ejercicio 12.

La chef en jefe de cierto restaurante acaba de recibir dos docenas de jitomate de su proveedor, pero todavía no los acepta. Sabe por la factura que el peso promedio de un jitomate es 7.5 onzas, pero insiste en que todos tengan un peso uniforme. Aceptará los jitomates sólo si el peso promedio es 7.5 onzas y la desviación estándar es menor que 0.5 onzas. Los pesos de los jitomates son los siguientes:

6.3	7.2	7.3	8.1	7.8	6.8	7.5	7.8	7.2	7.5	8.1	8.2
8.0	7.4	7.6	7.7	7.6	7.4	7.5	8.4	7.4	7.6	6.2	7.4

¿Cuál es la decisión de la chef y por qué?

Ejercicio 13.

Los siguientes datos son una muestra de la tasa de producción diaria de botes de fibra de vidrio que fabrica cierta compañía.

17	21	18	27	17	21	20	22	18	23

El gerente de producción de la compañía siente que una desviación estándar de más de tres botes por día indica variaciones de tasas de producción inaceptables. ¿Deberá preocuparse por las tasas de producción de la planta?

Ejercicio 14.

Dada la siguiente distribución de edades, x, en años, de niños que han padecido varicela:

X	1	2	3	4	5	6	7	8	9	10	11
F	4	9	11	8	7	5	4	2	2	0	1

Calcula la desviación estándar.

Ejercicio 15.

El número de cheques cobrados diariamente en las cinco sucursales de cierto Banco durante el mes anterior tuvo la siguiente distribución de frecuencias.

Para aprender más

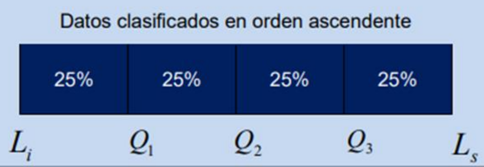
CUARTILES, DECILES Y PERCENTILES

Aunque la varianza y la desviación estándar son las medidas de dispersión más útiles en análisis estadístico, existen otras técnicas con las cuales puede medirse la dispersión de un conjunto de datos. Estas medidas adicionales de dispersión son los **cuartiles**, **los deciles** y los **percentiles**.

Cada conjunto de datos tiene tres cuartiles que lo dividen en cuatro partes iguales.

Cuartiles.

Son valores de la variable que dividen los datos ordenados en cuartos; cada conjunto de datos tiene tres cuartiles. El *primer cuartil*, Q_1 ,es un número tal que a lo sumo 25% de los datos son menores en valor que Q_1 y a lo sumo 75% son mayores. El segundo cuartil es la mediana (50%). El *tercer cuartil*, Q_3, es un número tal que a lo sumo 75% de los datos son menores en valor que Q_3 y a lo sumo 25% son mayores.



El primer cuartil es ese valor debajo del cual clasifica el 25% de las observaciones, y sobre el cual puede encontrarse el 75% restante. El segundo cuartil es justo la mitad. La mitad de las observaciones están por debajo y la mitad por encima; en este sentido, es lo mismo que la mediana. El tercer cuartil es el valor debajo del cual está el 75% de las observaciones y encima del cual puede encontrarse el 25% restante.

La determinación de los cuartiles con frecuencia es de utilidad. Por ejemplo, muchas escuelas de posgrados admitirán sólo a aquellos estudiantes que estén en el 25% superior (tercer cuartil) de los candidatos. Las empresas, con frecuencia, desean señalar las plantas cuyos deficientes registros de producción los colocan por debajo del cuartil inferior. Con un poco de imaginación es posible prever numerosos ejemplos en los cuales la determinación de cuartiles puede ser de gran beneficio.

deciles.

Son valores de la variable que dividen los datos ordenados en diez partes iguales (9 divisiones).

Datos clasificados en orden

L	D	D	D	D	D	D	D	D	D	L_s
10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	

Los es separan un conjunto de datos en 10 subconjuntos iguales, y los percentiles en 100 partes. El primer Décil es la observación debajo de la cual se encuentra el 10% de las observaciones, mientras que el 90% restante se encuentra encima de éste. El primer percentil es el valor debajo del cual se encuentra el 1% de las observaciones, y el resto están encima de éste. Puede aplicarse una interpretación similar al resto de es y percentiles. Todo conjunto de datos tiene 9 es y 99 percentiles.

Percentiles.

Son los valores de la variable que dividen un conjunto de datos clasificados en 100 subconjuntos iguales; cada conjunto de datos tiene 99 percentiles. El k-ésimo percentil, P_k , es un valor que a lo sumo k% de los datos son menores en valor que P_k y a lo sumo (100 - k)% de los datos son mayores.

Datos clasificados en orden ascendente

1%	1%	1%	1%	…	1%	1%	1%	1%

L_i P_1 P_2 P_3 P_4 P_{96} P_{97} P_{98} P_{99} L_s

Un percentil y su ubicación en un arreglo ordenado se identifica mediante los subíndices. Por ejemplo, el decimoquinto percentil se indica como P_{15}, y su ubicación en la serie ordenada es L_{15}.

Para ilustrar el cálculo de percentiles, se asume que se tienen observaciones para el número de acciones correspondientes a 50 acciones transados en la Bolsa de Valores de Nueva York, como se muestra en la siguiente tabla. Vale la pena destacar que los datos han sido puestos en una serie ordenada. El lugar del *P* ésimo percentil se halla

Ubicación de un percentil $$Lp = (n+1)\frac{P}{100}$$

En donde es el sitio del percentil en una serie ordenada
es el número de observaciones
es el percentil deseado

Se asume que se desea calcular el percentil 25, P_{25}, para las acciones de la tabla. Se debe hallar el primero su ubicación en la serie ordenada.

Números de acciones transadas en la Bolsa de Valores de Nueva York (en 100's)

3	10	19	27	34	38	48	56	67	74
4	12	20	29	34	39	48	59	67	74
7	14	21	31	36	43	52	62	69	76
9	15	25	31	37	45	53	63	72	79
10	17	27	34	38	47	56	64	73	80

El valor resultante de 12.75 decide que el percentil 25 está ubicado al 75% del trayecto comprendido entre la doceava observación, que es 20 y la treceava observación que es 21, es decir, P_{25} =20 + 0.75 (21-20) = 20.75.

Si se desea calcular el percentil 35, se halla

El percentil 35 está al 85% del trayecto comprendido entre la observación 17, que es 29 y la observación 18 que es 31, es decir $P_{35} = 29 + (0.85)(31-29) =$ 30.7. Por tanto, el 35% de las observaciones está por debajo de 30.7 y el 65% restante por encima de 30.7.

Regresando a los es y cuartiles por un momento, se nota que el primer Décil es igual a P_{10}, el segundo Décil es igual a P_{20}, y así sucesivamente. Adicionalmente, el primer cuartil es igual a P_{25}, el segundo cuartil es igual a P_{50}, y P_{75}, se encuentra en el tercer cuartil. Teniendo esto en mente, el cálculo de es y cuartiles se vuelve simplemente un asunto de determinación de los percentiles apropiados de acuerdo con las reglas que se acaban de establecer.

Ejemplo 1
Para la siguiente colección de datos 1, 1, 1, 2, 3, 3, 4, 4, 5, 9, 9, 19, 20 y 20 calcule:

a) El primero y el tercer cuartil. Ubíquelos en un diagrama de caja.
b) El octavo Décil.
c) El percentil 42, el 50 y el 87.

Solución Cuartiles

Para el cálculo de los cuartiles debemos determinar la posición del dato que ocupa cada cuartil con la condición de que dividan a la colección de datos en cuatro partes iguales. De esta forma encontramos con que el segundo cuartil coincide con la mediana dado que divide a la colección en dos partes iguales,

por lo que su posición es

$$n + 1 = \frac{14 + 1}{2} = \frac{7.5}{2}$$

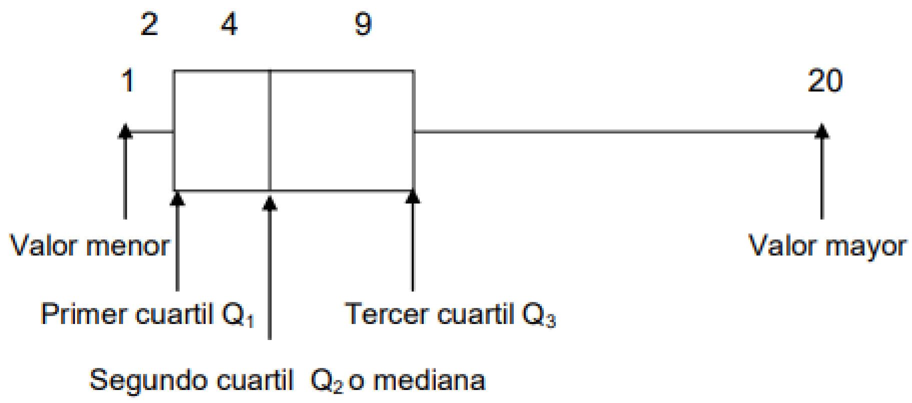
1 1 1 2 3 3 4 4 5 9 9 19 20 20

Posición 7.5 Valor de $Q_2 = 4$

Esto significa que la mediana es de 4 ahora el primer cuartil es la mediana de los datos que se encuentran a la izquierda de la mediana o segundo cuartil

Así la posición del primer cuartil es ${(n+1)}/{2} = {(7+1)}/{2} = 4$ Esto significa que el primer cuartil es el valor que está en la cuarta posición es decir 2. Del mismo modo, el tercer cuartil es el valor que está en la cuarta posición desde el otro extremo 9.

Como vemos en el diagrama de caja siguiente, al ubicar la caja entre el primero y el tercer cuartil, se puede tener una idea de la distribución de los datos, es decir, se observa que hay una mayor concentración de datos hacia los valores pequeños puesto que la caja está desplazada a la izquierda.



Valor menor Valor mayor

Primer cuartil Q_1 Tercer cuartil Q_3 Segundo cuartil Q_2 o mediana

deciles

En cuanto al octavo Décil, bastaría con ubicar la posición en que se encuentra a través de la fórmula 8 *n/10* = 8*14 /10 = 11.2. Esto quiere decir que entre el dato que se encuentra en la posición 11 Y LA 12

Está el octavo Décil, pero más cerca de la 11 que de la 12 puesto que la posición es la 11.2. El resultado sería **9.2** porque entre el 9 y el 19 (que son los datos cuyas posiciones son 11 y 12 respectivamente) hay exactamente 10 unidades.

Percentiles

Con relación a los percentiles perdidos, tendríamos que ubicar las posiciones correspondientes como lo hicimos con los es. Para la posición del percentil 42 tendríamos la siguiente fórmula $42 * \frac{n}{100} = 42 * \frac{14}{100}$=5.88 Esto quiere decir que el percentil 42 se encuentra entre los datos que ocupan la posición 5 y la 6. Afortunadamente en este caso ambos datos son 3 por lo que el percentil 42 es **3**. .88. Esto quiere decir que para el percentil 50 basta con buscar la mediana puesto que coinciden. La mediana de esta colección es 4. Por último, el percentil 87 se buscará con él. Mismo procedimiento usado anteriormente, es decir,

$$\frac{87n}{100} = \frac{87 * 14}{100} = 12.8$$

Nos daría la posición del percentil buscado que en este caso es entre las posiciones 12 y la 13, más cerca de la primera. El resultado sería que el percentil 87 toma el valor de **19.18**.

Ejemplo 2.

Ejemplo: En la siguiente serie simple, que corresponde a la edad de los trabajadores de una micro empresa: 33, 26, 66, 45, 28, 59, 33, 36, 26, 45, 62, 45, ordenar los datos y calcular los cuartiles uno, dos y tres, los es uno, tres, cinco y nueve; y, los percentiles nueve, diez y cincuenta.

Solución.

Ordenamos los datos de mayor a menor: 26, 26, 28, 33, 33, 36, 45, 45, 45, 59, 62, 66

Cuartiles

Hallamos la ubicación del cuartil uno con la fórmula:

$$Q_1 = \frac{n+1}{4}$$

$$Q_1 = \frac{12+1}{4} = 3.25$$

Calculamos el valor del cuartil uno:

El primer cuartil se localiza entre el tercer y cuarto valor y se encuentra a 0.25 de la distancia entre ellos. Como el tercer valor es 28, y el cuarto es 33, obtenemos la distancia entre ellos restando el valor mayor del menor; es decir, 33 – 28 = 5. Para ubicar el primer cuartil, hay que moverse a 0.25 de distancia entre el tercer valor y el cuarto, por lo que 0.25(5) = 1.25. Para terminar el procedimiento, sumamos 1.25

al primer valor, y resulta así que el primer cuartil es:

$Q_1 = 28 + 1.25 = 29.25$

Hallamos la ubicación del cuartil dos con la fórmula:

$$Q_2 = \frac{n+1}{2}$$

$$Q_2 = \frac{12+1}{2} = 6.5$$

Calculamos el valor del cuartil dos:

$$Q_2 = \frac{45+36}{2} = 40.5$$

Hallamos la ubicación del cuartil tres con la fórmula:

$$Q_3 = \frac{3(n+1)}{4}$$

$$Q_3 = \frac{3(12+1)}{4} = 9.75$$

Calculamos el valor del cuartil tres:

El tercer cuartil se localiza entre el noveno y décimo valor y se encuentra a 0.75 de la distancia entre ellos. Como el noveno valor es 45, y el décimo es 59, obtenemos la distancia entre ellos restando el valor mayor del menor; es decir, 59 – 45 = 14. Para ubicar el tercer cuartil, hay que moverse a 0.75 de distancia entre el noveno valor y el décimo, por lo que 0.75(14) = 10.5. Para terminar el procedimiento, sumamos 10.5 al primer valor, y resulta así que el tercer cuartil es:

$Q_3 = 45 + 10.5 = 55.5$

Hallamos la ubicación del Décil uno con la fórmula:

$$D_1 = \frac{n+1}{10}$$

$$D_1 = \frac{12+1}{10} = 1.3$$

Calculamos el valor del Décil uno:

El primer Décil se localiza entre el primero y segundo valor y se encuentra a 0.3 de la distancia entre ellos. Como el primer valor es 26, y el segundo es 26, se asume que el valor del primer Décil es de 26.

$D_1 = 26$

Hallamos la ubicación del tres con la fórmula:

$$D_3 = \frac{3(n+1)}{10}$$

$$D_3 = \frac{3(12+1)}{10} = 3.9$$

Calculamos el valor del tres:

El tercer se localiza entre el tercer y cuarto valor y se encuentra a 0.9 de la distancia entre ellos. Como el tercer valor es 28, y el cuarto es 33, obtenemos la distancia entre ellos restando el valor mayor del menor; es decir, 33 – 28 = 5. Para ubicar el tercer, hay que moverse a 0.9 de distancia entre el tercer valor y el cuarto, por lo que 0.9(5) = 4.5. Para terminar el procedimiento, sumamos 4.5 al primer valor, y resulta así que el tercer es: $D_3 = 28 + 4.5 = 32.5$

Hallamos la ubicación del cinco con la fórmula:

$$D_5 = \frac{n+1}{2}$$

$$D_5 = \frac{12+1}{2} = 6.5$$

Calculamos el valor del cinco:

$$D_5 = \frac{45+36}{2} = 40.5$$

Hallamos la ubicación del nueve con la fórmula:

$$D_9 = \frac{9(n+1)}{10}$$

$$D_9 = \frac{9(12+1)}{10} = 11.7$$

Calculamos el valor del nueve:

El noveno se localiza entre el onceavo y doceavo valor y se encuentra a 0.7 de la distancia entre ellos. Como el onceavo valor es 62, y el doceavo es 66, obtenemos la distancia entre ellos restando el valor mayor del menor; es decir, 66 – 62 = 4. Para ubicar el noveno, hay que moverse a 0.7 de distancia entre el onceavo valor y el doceavo, por lo que 0.7(4) = 2.8. Para terminar el procedimiento, sumamos 2.8 al primer valor, y resulta así que el noveno es:

$D_9 = 62 + 2.8 = 64.8$

Percentiles

Hallamos la ubicación del percentil diez con la fórmula:

$$P_{10} = \frac{n+1}{10}$$

$$P_{10} = \frac{12+1}{10} = 1.3$$

Calculamos el valor del percentil diez:

Por lo tanto, el valor de $P_{10} = 26$

Hallamos la ubicación del percentil cincuenta con la fórmula:

$$P_{50} = \frac{n+1}{2}$$

$$P_{50} = \frac{12+1}{2} = 6.5$$

Calculamos el valor del percentil cincuenta:
Por lo tanto, el valor de $P_{50} = 40.5$

Hallamos la ubicación del percentil noventa con la fórmula:

$$P_{90} = \frac{9(n+1)}{10}$$

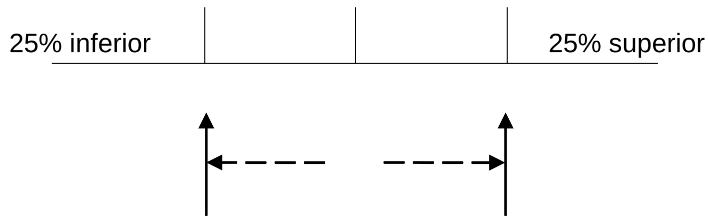
$$P_{90} = \frac{9(12+1)}{10} = 11.7$$

Calculamos el valor del percentil noventa:

$$P_{90} = 62 + 2.8 = 64.8$$

Una medida única de dispersión es el **rango o recorrido intercuartílico (interquartile range – RIQ).** EL RIQ es la diferencia entre el tercer cuartil y el primer cuartil. Es decir, $P_{75} - P_{25}$. La mitad de las observaciones se clasifican dentro de este rango. Consta del 50% de la mitad de las observaciones y corta el 25% inferior y el 25% superior de los puntos de datos. Como resultado, el RIQ proporciona una medida de dispersión que no está muy influenciada por unas cuantas observaciones extremas. El rango intercuartil se ilustra en la figura siguiente

Recorrido intercuartílico



RIQ

50% centrado

Ejercitando mi habilidad. (Producto esperado)

<u>**Ejercicios de cuartiles, deciles, percentiles y varianza**</u>

1.- Una distribución estadística viene dada por la siguiente tabla:

	[10, 15)	[15, 20)	[20, 25)	[25, 30)	[30, 35)
f_i	3	5	7	4	2

Hallar:

a) Varianza.

b) Los **cuartiles** 1º y 3º.

c) Los decil**es** 3º y 6º.

d) Los **percentiles** 30 y 70.

2. Dada la distribución estadística:

	[0, 5)	[5, 10)	[10, 15)	[15, 20)	[20, 25)	[25, ∞)
f_i	3	5	7	8	2	6

Calcular:

a) La **varianza**.

b) Cuartil 2º y 3º.

c) El 1°, 3° y 8°

3. El histograma de la distribución correspondiente al peso de 100 alumnos de Bachillerato es el siguiente:

a. Formar la **tabla de la distribución**.

b. Si Andrés pesa 72 kg, ¿cuántos alumnos hay menos pesados que él?

c. Hallar la **mediana**.

d. ¿A partir de que valores se encuentran el **25%** de los alumnos más pesados?

e. Varianza.

f. Desviación.

g. Coeficiente de Variación.

h. Es simétrica?

i. Cuál es el coeficiente de curtosis?

8. De esta **distribución de frecuencias absolutas acumuladas**, calcular:

Edad	F_i
[0, 2)	4
[2, 4)	11
[4, 6)	24
[6, 8)	34
[8, 10)	40

a) Varianza.
b) Desviación.
c) Coeficiente de Variación.
d) Es la distribución ¿simétrica?
e) ¿Cuál es el coeficiente de curtosis?

Para aprender más

ASIMETRÍA O SESGO

Así se le denomina al grado que tiende la distribución de frecuencias a sesgarse con relación al eje central (media aritmética).

a) Asimetría hacia la izquierda.
b) Simetría.
c) Asimetría hacia la derecha.

En épocas antiguas el signo que representaba a la Asimetría era el símbolo $\propto$ y este proporciona la posición de la media aritmética respecto a la moda, y, la fórmula de la Asimetría absoluta se presenta de la siguiente manera:

$\pm\propto = X - Mo$

Todo ello dependiendo de qué X sea $<$ que la Mo.

Se recomendaba por las limitantes de esta fórmula la Asimetría Relativa la cual se representaba así:

$$\pm IA = (X - Mo) / S$$

Ante el problema de tener dos modas Pearson demostró que la mediana ubica a dos segmentos respecto a la moda y uno respecto a la media aritmética, eliminando con esto el problema de la moda (bimodal) quedando la fórmula de la forma siguiente:

1. Índice de Asimetría de Pearson

$$IA = \frac{3(X - Md)}{S}$$

Existen dos métodos más para poder determinar la Asimetría de una distribución y son las siguientes:

2. Índice de Asimetría Por Momentos

$$A^3 = \frac{m^3}{S^3}$$

3. Índice de Asimetría Cuartílica

$$IQ = \frac{Q3 - 2Q2 + Q1}{Q3 - Q1}$$

ESCALA DE ASIMETRÍA

- Pearson y Momentos

			/////////	/////////			
INTENSA	ACENTUADA	MODERADA	DEBIL	DEBIL	MODERADA	ACENTUADA	INTENSA
-3 -2.25	-1.50	-0.60	0	0.60	1.50	2.25	3
			NORMAL				

- Cuartílica

			NORMAL				
-1	-0.90	-0.50	-0.10	0	0.10	0.50	0.90 1
			/////////	/////////			
NEGATIVA				POSITIVA			

EJEMPLO

i	X´i-1 - X´i	Xi	ni	Nj	xini	ni(xi - x)^2	Ui	niui	ni(ui)^2	ni(ui)^3	ni(ui)^4
1	60 - 62	61	5	5	305	208.01	0	0	0	0	0
2	63 - 65	64	18	23	1152	214.25	1	18	18	18	18
3	66 - 68	67	42	65	2814	8.51	2	84	168	336	672
4	69 - 71	70	27	92	1890	175.57	3	81	243	729	2187
5	72 - 74	73	8	100	584	246.42	4	32	128	512	2048

Σ	**100**	**6745**	**852.75**	**215**	**557**	**1595**	**4925**

I. ASIMETRIA DE PEARSON

$$IA = \frac{3(X - Md)}{S}$$

$$X = \frac{\Sigma X i n i}{N} = \frac{6,475}{100} = 67.45$$

$$Md = X`_{j-1R} + \frac{(N/2 - Nj\text{-}1)}{nj} C$$

$$Md = 65.5 + \frac{(50\text{-}23)}{} 3 = \frac{(50-23)}{42} \; 67.43$$

$$S = \frac{\sqrt{\sum ni(x_i - x)^2}}{N} = \frac{\sqrt{852.76}}{100} = 2.92$$

$$IA = \frac{3(67.45 - 67.43)}{2.92} = 0.02$$

Normal Débil

II. ASIMETRIA CUARTILICA

$$IQ = \frac{Q3 - 2Q2 + Q1}{Q3 - Q1}$$

$$Q_1 = X`_{j-1R} + \frac{(1/4N - Nj\text{-}1)}{Nj} C$$

$$Q_1 = 65.5 + \frac{(25\text{-}23)}{42} 3 = 65.64$$

$$Q_3 = X`_{j-1R} + \frac{(3/4N - Nj\text{-}1)}{nj} C$$

$Q_3 = 68.5 + \frac{(75-65)}{27} 3 = 69.61$

$IQ = \frac{69.61 - 2(67.43) + 65.64}{69.61 - 65.64} = 0.10$ **Normal**

Positiva

III. ASIMETRIA POR MOMENTOS

$A3 = \frac{m3}{S^3}$

Momentos Simples

$m`1 = \frac{(\Sigma niui)}{N} C$ $= \frac{(215)}{100} 3 = 6.45$

$m`2 = \frac{(\Sigma niui\text{^}2)}{N} C\text{^}2$ $= \frac{(557)}{100} 3\text{^}2 = 50.13$

$m`3 = \frac{(\Sigma niui\text{^}3)}{N} C\text{^}3$ $= \frac{(1595)}{100} 3\text{^}3 = 430.65$

$m`4 = \frac{(\Sigma niui\text{^}4)}{N} C\text{^}4$ $= \frac{(4925)}{100} 3\text{^}4 = 3989.25$

Momentos respecto a la media

m1 = 0

m2 = m`2 – (m`1)^2

m2 = 50.13 – (6.45)^2 = 8.53

m3 = m`3 – 3(m`1 *m`2)+2(m`1)^3

m3 =430.65–3(6.45 *50.13)+2(6.45)^3=-2.69

m4 =m`4–4(m`1*m`3)+6(m`1)^2(m`2)-3(m`1)^4

m4=3989.25-4(6.45*430.65)+6(6.45)^2(50.13)-3(6.45)^4 = 199.38

$$A^3 = \frac{m^3}{S^3} = \frac{-2.69}{2.92^3} = -0.11$$

Normal Débil

Fuente: **https://www.academia.edu/search?utf8=%E2%9C%93&q=SESGO+**

Ejercitando mi habilidad.

Los siguientes valores corresponden a la temperatura máxima diaria (ºF) de 36 días, obtenidos a las 14 horas en una cierta estación meteorológica.

84, 49, 61, 40, 83, 67, 45, 66, 70, 69, 80, 58, 68, 60, 67, 72, 75, 76, 73, 70, 63, 70, 78, 52, 67, 53, 67, 75, 61, 70, 81, 76, 79, 58, 57, 21

. a) Calcular: media, desviación típica muestral, cuartiles superior e inferior y la mediana.

b) Estudiar la existencia de datos atípicos. Si existe algún valor atípico omitir, dicho valor y calcular de nuevo el apartado a).

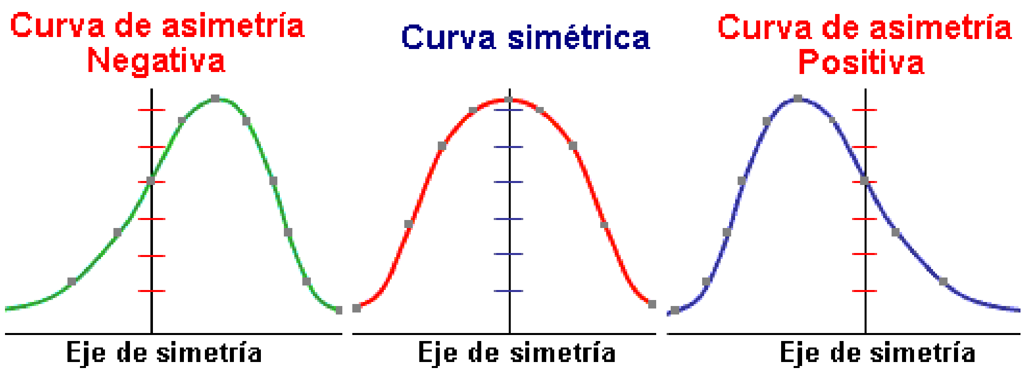
c) Con los datos de los apartados a y b construir un gráfico con el diagrama de caja, de ambos apartados.

Para aprender más

MEDIDAS DE ASIMETRIA Y CURTOSIS

MEDIDAS DE ASIMETRIA

Las medidas de asimetría son aquellas que nos muestran el grado de deformación o sesgo de una distribución de frecuencias. Una distribución puede presentar una de las siguientes formas:



Se dice que una distribución es asimétrica derecha cuando esta sesgada a la derecha, asimismo se dice que una distribución es asimétrica izquierda cuando esta sesgada a la izquierda. Una distribución se dice simétrica cuando no presenta sesgo alguno.

Entre las principales medidas descriptivas para medir el grado de asimetría de una distribución se tiene al coeficiente de asimetría de Karl Pearson, asimetría de Bowley, y la simetría por momentos.

ASIMETRIA DE KARL PEARSON

$$As = \frac{\bar{X} - Mo}{S_x}$$

ASIMETRIA POR LA RELACION EMPÍRICA

$$As = \frac{3(\bar{X} - Me)}{S_x}$$

ASIMETRIA POR MOMENTOS

$$b_1 = \frac{M_3}{(S_x)^3}$$

Fórmula para una serie simple:

$$b_1 = \frac{\dfrac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^3}{n}}{(S_x)^3}$$

Fórmula para una serie agrupada:

$$b_1 = \frac{\dfrac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^3 f_i}{n}}{(S_x)^3}$$

En cualquiera de los casos si:

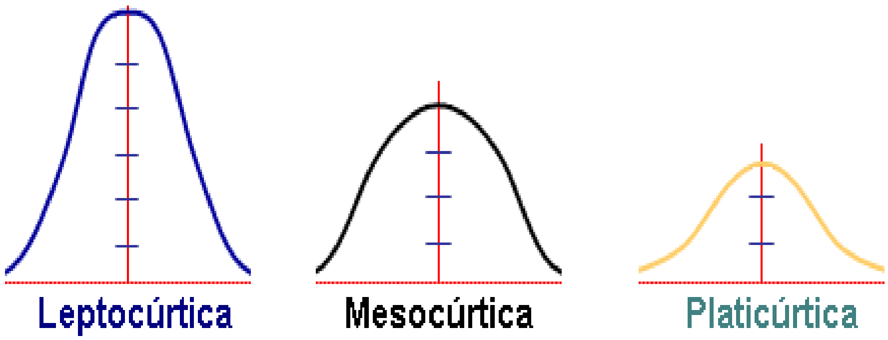
As > 0, la distribución es asimétrica derecha

As = 0, La distribución es simétrica

As < 0, la distribución es asimétrica izquierda

MEDIDAS DE CUSTORSIS

La curtosis nos muestra el grado de apuntamiento o concentración de una distribución con respecto a media aritmética. En una distribución se pueden tres tipos de distribuciones como se muestra en la figura:



Se dice que una distribución es leptocúrtica cuando esta es apuntada lo cual nos mostrará un alto grado de concentración de los valores de la variable. Asimismo se dice que una distribución es platicúrtica cuando esta es achatada mostrando de este modo un alto grado de dispersión de los valores de la variable. Luego, una distribución se dice mesocúrtica cuando es no es ni apuntada ni achatada, es decir es semejante a una distribución normal.

El estadígrafo utilizado para medir la curtosis de una distribución está definida por la relación entre el cuarto momento central y la desviación típica a la cuarta, cuya fórmula es:

$$b_2 = \frac{M_4}{(S_x)^4}$$

Fórmula para una serie simple

$$b_2 = \frac{\dfrac{\sum_{i=1}^{n} \left(X_i - \bar{X} \right)^4}{n}}{(S_x)^4}$$

Fórmula para una serie agrupada

$$b_2 = \frac{\dfrac{\sum_{i=1}^{n} \left(X_i - \bar{X} \right)^4 f_i}{n}}{(S_x)^4}$$

El valor del coeficiente de curtosis se compara con el valor de distribución normal que es igual a K = 3 de modo que:

$b_2 > k :$ La distribución es leptocúrtica

$b_2 < k :$ La distribución es platicúrtica

$b_2 = k :$ La distribución es mesocúrtica

MOMENTOS

Se llaman así a un conjunto de promedios que engloban algunos conceptos referidas a las medidas descriptivas como la media, la varianza, etc.

Los momentos pueden ser respecto al origen y respecto de un a medida de posición, siendo este último momentos centrales

- MOMENTOS
- AL ORIGEN
- CENTRALES

Los momentos al origen se definen como el promedio de los valores de una variable elevadas a una potencia enésima.

Los momentos centrales se definen como el promedio de los desvíos de una variable respecto de la media elevadas a una potencia enésima

Fórmula para una serie simple:

$$a_r = \frac{\sum_{i=1}^{n} X^r}{n} \qquad m_r = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^r}{n}$$

Fórmula para una serie agrupada:

$$a_r = \frac{\sum_{i=1}^{n} X^r f_i}{n} \qquad m_r = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^r f_i}{n}$$

Donde r = 1,2, 3….

OTRAS MEDIDAS DE CENTRALIZACIÓN

MEDIA GEOMETRICA

La media geométrica se define como la raíz enésima del producto de los valores de la variable para una serie simple, y como la raíz enésima del producto de los valores de la variable elevadas a sus respectivas frecuencias para una serie agrupada . Se simboliza por “Mg” y su fórmula es:

Fórmula para una serie simple

$$Mg = \sqrt[n]{x_1 x_2 x_3 \ldots\ldots\ldots x_n}$$

Fórmula para una serie agrupada.

$$Mg = \sqrt[n]{x_1^{f_1} x_2^{f_2} x_3^{f_3} \ldots\ldots\ldots\ldots x_n^{f_n}}$$

MEDIA ARMONICA

La media armónica se define como la relación del total de observaciones sobre la suma de los valores inversos de una variable. Es la inversa de la media aritmética y se simboliza por “H”, así:

Fórmula para una serie simple:

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \ldots\ldots + \frac{1}{x_{n1}} +} = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}}$$

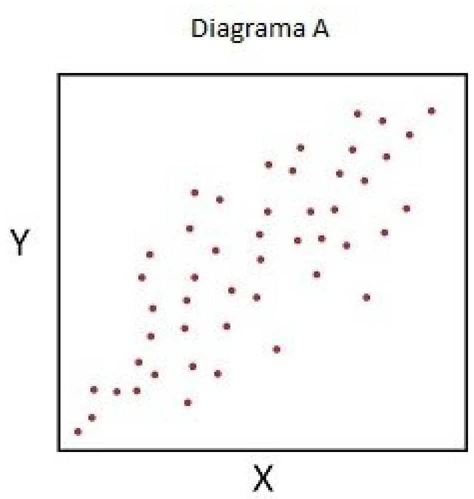
Fórmula para una serie agrupada:

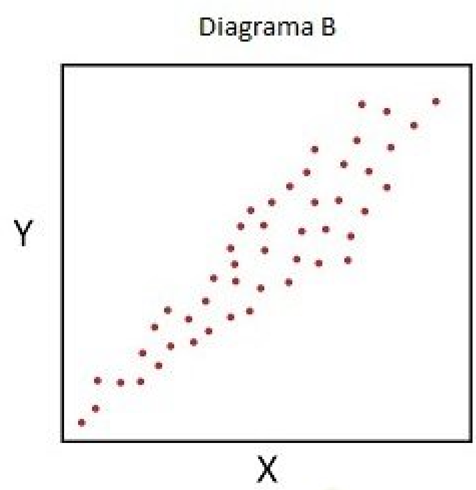
$$H = \frac{n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \frac{f_3}{x_3} + \ldots\ldots\ldots + \frac{f_n}{x_{n1}} +} = \frac{n}{\sum_{i=1}^{n} \frac{f_i}{X_i}}$$

EL COEFICIENTE DE CORRELACIÓN

El Coeficiente de correlación es una medida que permite conocer el grado de asociación lineal entre dos variables cuantitativas (X, Y).

En los siguientes Diagramas de dispersión se puede observar que existe una relación lineal entre la variable X y la variable Y.





Sin embargo, si trazamos una línea recta en los diagramas.

Diagrama A

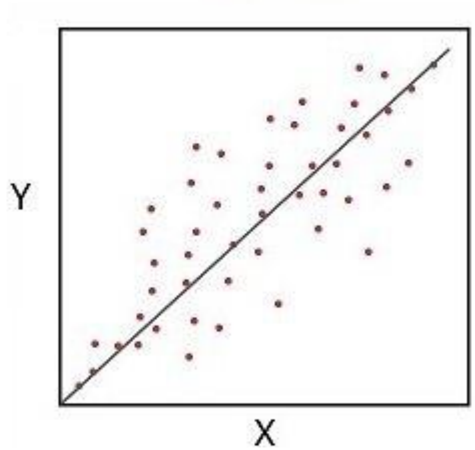
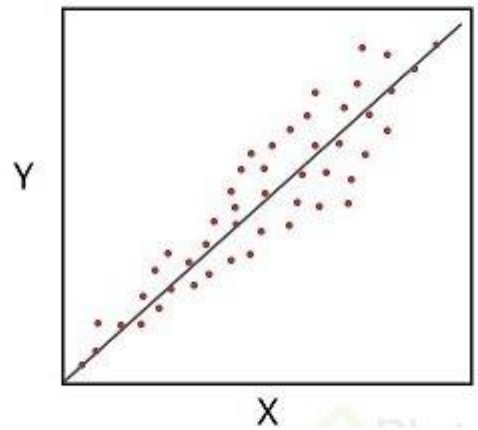


Diagrama B



Podemos observar que en un diagrama B los puntos se acercan más a la recta, caso contrario en el diagrama A, los puntos están más alejados. Entonces podemos decir que la relación lineal del diagrama A es más débil con comparación a la relación que existe en el diagrama B.

Un diagrama dispersión no nos da certeza de que tan débil o fuerte es la relación lineal, necesitamos una medida que nos de la fuerza de la asociación y la dirección que toma esta relación.

Para esto sirve el coeficiente de correlación que está dado por la siguiente formula.

$$r = \frac{S_{XY}}{S_X \, S_Y}$$

¿De dónde sacamos estos valores?

S_{XY} = Covarianza

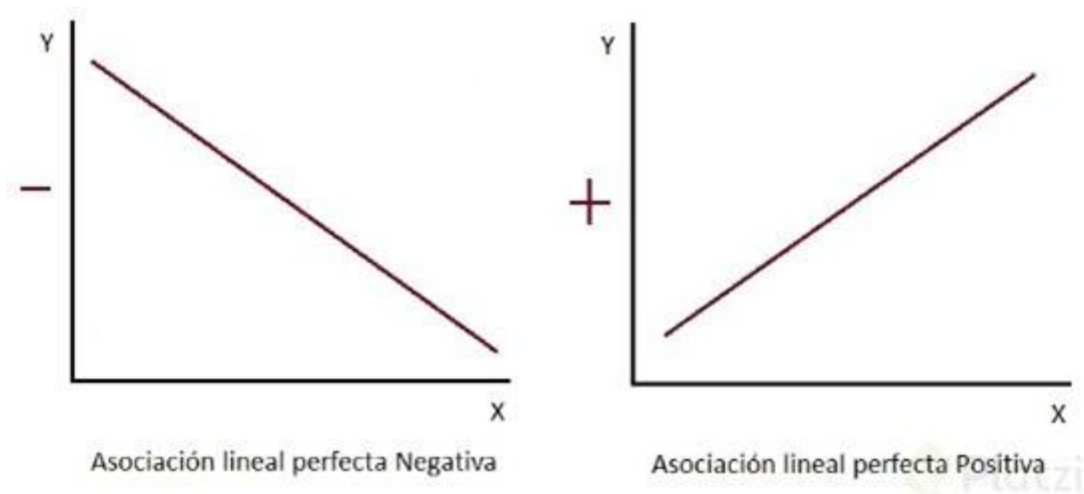
$S_X \, S_Y$ = Desviacion Estandar de X multiplicada por la Desviacion Estandar de Y

Recordar entonces que el coeficiente de relación lineal, mide la fuerza y el sentido de la relación lineal entre 2 variables cuantitativas.

Luego de haber aplicado la fórmula, según el resultado se puede clasificar en este rango.

$$r_{\text{(coeficiente de correlación)}} = \left[\begin{array}{c} -1 \\ 0 \\ 1 \end{array}\right.$$

Entre más cercano es a 1 es más fuerte, entre más cercano a 0 es débil hasta llegar hacerse nula, si los valores del coeficiente de relación son -1 es una Asociación lineal perfecta Negativa, si es 0 no existe relación y si es 1 es una Asociación Lineal perfecta Positiva.



¿Entonces cómo identificamos cuan dispersos son los puntos de la línea recta?

Con la siguiente tabla podemos clasificar nuestros resultados y responder esta interrogante.

Rango		Relacion Lineal
±0,96	±1,0	Perfecta
±0,85	±0,95	Fuerte
±0,70	±0,84	Significativa
±0,50	±0,69	Moderada
±0,20	±0,49	Débil
±0,10	±0,19	Muy Débil
±0,09	±0,0	Nula

Ejercitando mi habilidad.

Instrucciones. – Resuelve los problemas que a continuación se enlistan

Una compañía desea hacer predicciones del valor anual de sus ventas totales en cierto país a partir de la relación de éstas y la renta nacional. Para investigar la relación cuenta con los siguientes datos:

X	Y
189	402
190	404
208	412
227	425
239	429
252	436
257	440
274	447
293	458
308	469
316	469

X representa la renta nacional en millones de pesos e Y representa las ventas de la compañía en miles de pesos en el periodo que va desde ___ hasta ___ (ambos inclusive). Calcular:

a) La recta de regresión de Y sobre X.
b) El coeficiente de correlación lineal e interpretarlo.
c) Si en la renta nacional del país fue de ___ millones de pesos. ¿Cuál será la predicción para las ventas de la compañía en este año?

1. La información estadística obtenida de una muestra de tamaño 12 sobre la relación existente entre la inversión realizada y el rendimiento obtenido en cientos de miles de pesos para explotaciones agrícolas, se muestra en el siguiente cuadro: Inversión (X), Rendimiento (Y)

X	Y
11	2
14	3
16	5
15	6
16	5
18	3
20	7
21	10
14	6
20	10
19	5
11	6

Calcular:

a) La recta de regresión del rendimiento respecto de la inversión.

b).-La previsión de inversión que se obtendrá con un rendimiento de 1 250 000 pesos.

El número de horas dedicadas al estudio de una asignatura y la calificación obtenida en el examen correspondiente, de ocho personas es: Horas (X)Calificación (Y)

X	Y
20	6.5
16	6
34	8.5
23	7
27	9
32	9.5
18	7.5
22	8

Se pide:

a) Recta de regresión de Y sobre X.
b) Calificación estimada para una persona que hubiese estudiado en horas.

2. En la tabla siguiente se indica la edad (en años) y la conducta agresiva (medida en una escala de cero a 10) de 10 niños. Edad Conducta Agresiva

Edad	Conducta agresiva
6	9
6	6
6.7	7
7	8
7.4	7
7.9	4
8	2
8.2	3
8.5	2
8.9	1

a) Obtener la recta de regresión de la conducta agresiva en función de la edad.

A partir de dicha recta, obtener el valor de la conducta agresiva que correspondería a un niño de años.

La recta de regresión de “x” sobre “y” se utiliza para estimar los valores de “y” a partir de los de “x”.

La pendiente de la recta es el cociente entre la covarianza y la varianza de la variable .

Si la correlación es nula, esto es, las rectas de regresión son perpendiculares entre sí.

Ejemplo de recta de regresión

1.-Las notas de 12 alumnos de una clase en Matemáticas y Física son las siguientes:

Matemáticas	2	3	4	4	5	6	6	7	7	8	10	10
Física	1	3	2	4	4	4	6	4	6	7	9	10

Hallar las rectas de regresión y representarlas

x_i	2	3	4	4	5	6	6	7	7	8	10	10	72
y_i	1	3	2	4	4	4	6	4	6	7	9	10	60
$x_i - y_i$	2	9	8	16	20	24	36	28	42	56	90	100	431
x_i^2	4	9	16	16	25	36	36	49	49	64	100	100	504
y_i^2	1	9	4	16	16	16	36	16	36	42	81	100	380

1. Hallamos las medias aritméticas.

$$\overline{x} = \frac{72}{12} = 6, \qquad \overline{y} = \frac{60}{12} = 5$$

2. Calculamos la covarianza

$$s_{xy} = \frac{1}{12} \cdot 431 - 6 \cdot 5 = 5.92$$

3. Calculamos las varianzas.

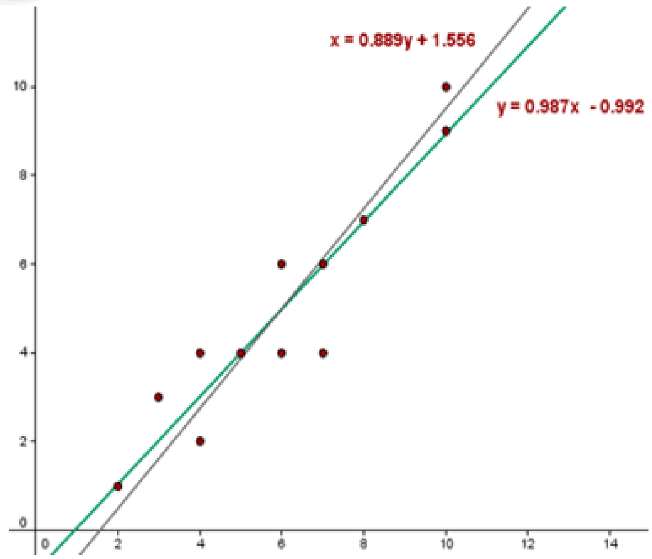
$$s_x^2 = \frac{1}{12} \cdot 504 - 6^2 = 6, \qquad s_y^2 = \frac{1}{12} \cdot 380 - 5 \cdot 5 = 6.66$$

4. Recta de regresión de “x” sobre “y” .

$$y - 5 = \frac{5.92}{6}(x - 6) \quad \Longrightarrow \quad y = 0.987x - 0.922$$

5. Recta de regresión de “y” sobre “x” .

$$x - 6 = \frac{5.92}{6.66}(y - 5) \quad \Longrightarrow \quad x = 0.889y + 1.556$$



Ejercitando mi habilidad.

Instrucciones. – Realiza y calcula lo que se pida en los siguientes ejercicios:

1. Cinco niños de 2, 3, 5, 7 y 8 años de edad pesan, respectivamente, 14, 20, 32, 42 y 44 kilos.

a) Hallar la ecuación de la recta de regresión de la edad sobre el peso.

b) ¿Cuál sería el peso aproximado de un niño de seis años?

2.- Las notas obtenidas por cinco alumnos en Matemáticas y Química son:

Matemáticas	Química
6	6.5
4	4.5
8	7
5	5
3.5	4

a) Determinar las rectas de regresión y calcular la nota esperada en Química para un alumno que tiene 7.5 en Matemáticas.

3.- Las estaturas y pesos de jugadores de baloncesto de un equipo son

Estatura (cm)	Pesos (kg)
186	85
189	85
190	86
192	90
193	87
193	91
198	93
201	103
203	100
205	101

Calcular:

a) La recta de regresión de altura sobre peso.
b) El coeficiente de correlación.
c) El peso estimado de un jugador que mide 195 cm.

Error Estándar

Error de tipo I. Se comete cuando la hipótesis nula es verdadera y, como consecuencia del contraste, se rechaza.

Error de tipo II. Se comete cuando la hipótesis nula es falsa y, como consecuencia del contraste se acepta.

H_0	Verdadera	Falsa
Aceptar	Decisión correcta Probabilidad = 1 - α	Decisión incorrecta: ERROR DE TIPO II
Rechazar	ERROR DE TIPO I Probabilidad = α	Decisión correcta

La probabilidad de cometer Error de tipo I es el nivel de significación α.

La probabilidad de cometer Error de tipo II depende del verdadero valor del parámetro. Se hace tanto menor cuanto mayor sea n.

Ejemplo.

Las ventas mensuales de una tienda de electrodomésticos se distribuyen según una ley normal, con desviación típica 900 pesos. En un estudio estadístico de las ventas realizadas en los últimos nueve meses, se ha encontrado un intervalo de confianza para la media mensual de las ventas, cuyos extremos son 4 663 pesos y 5 839 pesos.

a) ¿Cuál ha sido la media de las ventas en estos nueve meses?

b) ¿Cuál es el nivel de confianza para este intervalo?

c) ¿Cuál ha sido la media de las ventas en estos nueve meses?

$$n = 9 \quad x = (4663+5839) / 2;$$

$$x=5251$$

B.- ¿Cuál es el nivel de confianza para este intervalo?

$$E = (5839 - 4663) / 2 = 588$$

$$588 = z_{\alpha/2} \cdot 900 / 3 \qquad z_{\alpha/2} = 1.96$$

$$1-\alpha = 0.95 \quad \rightarrow \quad 95\%$$

Ejercitando mi habilidad.

Instrucciones. - Realiza los ejercicios como se te pide

1.- Se desea estimar la proporción, p, de individuos daltónicos de una población a través del porcentaje observado en una muestra aleatoria de individuos, de tamaño n.

a) Si el porcentaje de individuos daltónicos en la muestra es igual al 30%, calcula el valor de n para que, con un nivel de confianza de 0,95, el error cometido en la estimación sea inferior al 3,1%.

b) Si el tamaño de la muestra es de 64 individuos, y el porcentaje de individuos daltónicos en la muestra es del 35%, determina, usando un nivel de significación del 1%, el correspondiente intervalo de confianza para la proporción de daltónicos de la población.

2.- Una marca de nueces afirma que, como máximo, el 6% de las nueces están vacías. Se eligieron 300 nueces al azar y se detectaron 21 vacías.

a) Con un nivel de significación del 1%, ¿se puede aceptar la afirmación de la marca?

b) Si se mantiene el porcentaje muestral de nueces que están vacías y $1-\alpha = 0.95$, ¿qué tamaño muestral se necesitaría para estimar la proporción de nueces con un error menor del 1% por ciento?

3.- La duración de la bombilla de 100 W que fabrica una empresa sigue una distribución normal con una desviación típica de 120 horas de duración. Su vida media está garantizada durante un mínimo de 800 horas. Se escoge al azar una muestra de 50 bombillas de un lote y, después de comprobarlas, se obtiene una vida media de 750 horas. Con un nivel de significación de 0,01, ¿habría que rechazar el lote por no cumplir la garantía?

4.- Según el dueño de un taller mecánico sus mecánicos están reparando máximo 17 motos en promedio semanalmente para probarlo selecciono aleatoriamente 33 mecánicos y obtuvo el número de reparaciones hechas por cada uno durante una semana aleatoria. Los resultados arrojaron una media de 19. De un estudio previo se sabe que la varianza semanal en la cantidad de motos reparadas es de 10(motos al cuadrado).

Con base en la evidencia estadística y al 1% de significancia concluye conforme a la afirmación del dueño del taller

REFERENCIAS BIBLIOGRÁFICAS

- Probabilidad y estadísticas, Martha Patricia Sandoval, 2013.
- https://matematicasies.com/?page=mot_ej&id_mot=118
- Webster, Allen L., (2000)., *Estadística aplicada a los negocios y la economía*., Editorial McGraw-Hill., Colombia.
- http://www.iesxunqueira1.com/Download/pdf/ejdescriptiva%201.pdf
- http://translate.google.com.gt/translate?hl=es&sl=en&u=http://en.wikipedia.org/wiki/Percentile&ei=CZCqT5KDE4qFtgetvICiAg&sa=X&oi=translate&ct=result&resnum=3&ved=0CDkQ7gEwAg&prev=/search%3Fq%3Dpercentiles%26hl%3Des%26safe%3Doff%26biw%3D1024%26bih%3D667%26prmd%3Dimvns
- http://es.wikipedia.org/wiki/Medidas_de_posici%C3%B3n_no_central
- Las imágenes fueron tomadas de pixabay en noviembre 2020

PROBABILIDAD Y ESTADÍSITCA