Guía Didáctica
Estadística

Prof. Gabriel Castellano

Caracas, mayo 2018

ÍNDICE

INTRODUCCIÓN

La estadística es una herramienta fundamental en cualquier campo del saber y muy especialmente para los alumnos que se preparan en el área financiera, ya que con ella pueden hacer proyecciones y estimaciones que serán necesarias para lograr el mejor aprovechamiento de los recursos.

Mediante la asignatura comprenderá conceptos importantes en el ámbito mercantil y financiero, lo cual le garantizará una sólida formación técnica.

Para afianzar conocimientos se irá suministrando ejercicios prácticos, y al final de algunas unidades, habrá autoevaluaciones que el participante, después del estudio concienzudo, estará en capacidad de resolver.

El IUDAG, aportando valor agregado, le suministra esta guía al estudiante para coadyuvar en su acertada preparación.

Competencias que se espera alcanzar en el estudiante.

General

Aplica los conceptos, técnicas y procedimientos para el procesamiento, presentación y análisis de datos de la Estadística aplicada.

Específicas

1.- Define conceptos fundamentales de la estadística aplicada.
2.- Procesa datos recolectados por muestra, los presenta y analiza con fines de toma de decisiones y para presentación de informes.
3.- Aplica procedimientos para la construcción y análisis de tablas de distribución de frecuencia.
4.- Determina las medidas de tendencia central y las medidas de dispersión.
5.- Participa en la elaboración de modelos de regresión lineal y serie de tiempo.
6.- Analiza cuadros, gráficos y tablas estadísticas

FUNDAMENTOS

ESTADÍSTICA

Ciencia rama de la Matemática que se ocupa de recolectar, organizar, presentar, analizar e interpretar información cuantitativa para obtener conclusiones válidas, solucionar problemas, predecir fenómenos y ayudar a una toma de decisiones más efectivas.

APLICACIONES

Antes sólo se aplicaba a los asuntos del Estado, pero en la actualidad la utilizan las compañías de seguros, empresarios, comerciantes, educadores, etc. No hay campo de la actividad humana que no requiera del auxilio de esta ciencia, así por ejemplo:

- El educador, mediante la estadística, podrá conocer si un estudiante lee muy bien o regular, si la asistencia es normal o irregular, si la estatura está en relación con la edad, media aritmética de rendimiento escolar en un período determinado, etc.
- El hombre de negocios realiza encuestas estadísticas para determinar la reacción de los consumidores frente a los actuales productos de la empresa y en el lanzamiento de los nuevos.
- El economista emplea una amplia gama de elementos estadísticos para estudiar los planes de los consumidores y efectuar pronósticos sobre las tendencias de las actividades económicas
- El sociólogo trata de auscultar la opinión pública mediante encuestas, para determinar su preferencia por un candidato presidencial, o su posición frente a determinados problemas económicos, políticos o sociales
- El geólogo utiliza métodos estadísticos para determinar las edades de las rocas
- El Genetista determina las semejanzas entre los resultados observados y esperados en una experiencia genética, esto se determina estadísticamente

FINES

- **Conocer** las características de un grupo de casos de estudio.
- **Comparar** entre los resultados actuales y los obtenidos en experiencias pasadas para determinar las causas que han influenciado en los cambios.
- **Predecir** lo que pude ocurrir en el futuro de un fenómeno.

OBJETIVOS

- **Describir** numéricamente las características de los conjuntos de observaciones. Esta etapa consiste en recopilar, organizar, tabular y presentar gráficamente los datos, proporcionando una visión cuantitativa de los fenómenos observados.
- **Analizar** los datos de manera objetiva con el fin de disponer de un concepto claro de universo o población y adoptar decisiones basadas en la información proporcionada por los datos de la muestra.
- **Estimar** o predecir lo que sucederá en el futuro con un fenómeno de una manera relativamente aceptable, así por ejemplo, podemos estimar cuál será la población del país dentro de un determinado número de años conociendo la actual.

FASES DEL MÉTODO ESTADÍSTICO

- **Recopilación.-** Obtención de datos relacionados con el problema motivo de estudio, utilizando instrumentos, tales como: cuestionarios, entrevistas, informes, memorias, etc.
- **Organización.-** Se realizar crítica, corrección, clasificación y tabulación de los datos obtenidos en el paso anterior.
- **Presentación.-** Mostrar datos de manera significativa y descriptiva. Los datos deben colocarse en un orden lógico que revele rápida y fácilmente el mensaje que contienen. La presentación se la puede hacer a través de gráficos estadísticos.
- **Análisis.-** Descomponer el fenómeno en partes y luego examinar cada una de ellas con el objetivo de lograr una explicación, haciendo uso, en su mayoría, de los cálculos matemáticos.
- **Interpretación.-** Proceso mental, mediante el cual se encuentra un significado más amplio de los datos estadísticos con el objetivo de llegar a conclusiones para la toma de decisiones y solución de problemas.

CLASIFICACIÓN DE LA ESTADÍSTICA

- **ESTADÍSTICA DESCRIPTIVA O DEDUCTIVA**

  Es la parte de la Estadística que proporciona métodos para recopilar, organizar, presentar, resumir, analizar e interpretar la información contenida en un conjunto de datos, los cuales han de plasmarse en gráficos, tabulares o numéricos, así por ejemplo:

  Un docente calcula la calificación promedio de sus cursos. Sólo describe el desempeño, no hace ninguna generalización acerca de los mismos, acá se está haciendo uso de la Estadística Descriptiva.

- **ESTADÍSTICA INFERENCIAL O INDUCTIVA**

  Es la parte de la Estadística que proporciona métodos para extraer conclusiones o generalizaciones que sobrepasan los límites del conocimiento aportados por un conjunto de datos. Busca obtener información sobre la población, basándose en el estudio de los datos de una muestra tomada a partir de ella, así por ejemplo:

  El docente utiliza el promedio de calificaciones obtenidas por uno de sus cursos para estimar la calificación promedio de los 5 cursos a su cargo. Al realizar generalización acerca los diferentes cursos, en este caso se usa la Estadística Inferencial.

POBLACIÓN

Conjunto de todos los elementos que serán sometidos a un estudio estadístico, es decir, sobre el que se realizan las observaciones.

Una población puede ser finita o infinita.

- Finita cuando está delimitada y conocemos el número que la integran, así por ejemplo: Estudiantes del IUDAG.
- Infinita cuando a pesar de estar delimitada en el espacio, no se conoce el número de elementos que la integran, por ejemplo: Todos los profesionales universitarios que están ejerciendo su carrera, peces del mar, estrellas en el infinito, etc.

MUESTRA (N)

Es el subconjunto de una población, es un pequeño universo. Se la usa cuando la población es infinita o sumamente grande y es imposible Observar todos sus elementos.

Ejemplo: Estatura de los empleados de una fábrica. Calificaciones de los alumnos matriculados en Estadística en la Modalidad de Estudios a Distancia

VARIABLE

Caracteres susceptibles a cambio y pueden tener diferentes valores en cada elemento o individuo.

Clasificación de la variable

- ***Variable Cualitativa***

  Expresan distintas cualidades, características o modalidad. Cada modalidad que se presenta se denomina atributo o categoría, y la medición consiste en una clasificación de dichos atributos. Las variables cualitativas puede ser **dicotómicas** cuando sólo pueden tomar dos valores posibles, como *sí y no*, *hombre y mujer* o ser **Politómicas** cuando pueden adquirir tres o más valores.

  Dentro de ellas podemos distinguir:

  - ***Ordinal***

    Puede tomar valores ordenados siguiendo una escala establecida, por ejemplo: (grado de satisfacción: excelente, bueno, regular); (intensidad del dolor: leve, moderado, fuerte); (posición en la lista: 1°, “2°, 3°); etc.

  - ***Nominal***

    Los datos no pueden ser sometidos a un criterio de orden, por ejemplo: sexo, grupo sanguíneo, religión, nacionalidad, colores, etc.

- ***Variable Cuantitativa***

  Es toda magnitud representada por números. Como por ejemplo, peso, estatura, número de habitantes, etc. Se divide en:

- ***Variable Discreta***

  Está limitada a ciertos valores, generalmente números enteros (Z) o exactos, frecuentemente resultan de la enumeración o del conteo, como por ejemplo: número de estudiantes de la promoción, número de carros vendidos, etc. No existen valores fraccionados.

- ***Variable Continua***

  Puede tomar cualquier valor dentro de un intervalo, representado por un número racional (Q), por lo que su espacio muestral es infinito. Por ejemplo: (la masa: 2,3 kg, 2,4 kg, 2,5 kg…); (la altura: 1,64 m, 1,65 m, 1,66 m…)

INDIVIDUO O ELEMENTO

Unidad mínima que compone una población, es decir, cada uno de los integrantes

Puede ser

- Una entidad simple: una persona
- Una entidad compleja: una familia
- Algo con existencia real: un automóvil
- Algo abstracto: un voto, la temperatura, el tiempo
- Unidades naturales: obreros, turistas, empleados, emigrantes, etc.

PARÁMETRO

Conjunto de características (resultados) o valores numéricos, cuando se han obtenido a partir de una población. Ejemplo: Edad promedio de los alumnos de la Universidad.

ESTADÍSTICO

Conjunto de características (resultados) cuando se han obtenido a partir de una muestra.

Ejemplos básicos de parámetros estadísticos son: la media y la desviación estándar. Algunos ejemplos gráficos son: histograma, pirámide poblacional, clúster, entre otros.

DATOS ESTADÍSTICOS

Son medidas, valores o características susceptibles de ser observados y contados. Como por ejemplo, la edad de los estudiantes del IUDAG.

Los datos estadísticos pueden ser clasificados en:

- ***Cualitativos:*** la diferencia entre ellos es de clase y no de cantidad.
  Ejemplo: Estado civil, Sexo, Raza, Color, Nivel educativo
- ***Cuantitativos:*** representan magnitudes.
  Ejemplo***:*** Edad, Estatura, precio.
- ***Cronológicos:*** difieren en instantes o períodos de tiempo.
- ***Geográficos:*** referidos a localidad.

Los datos estadísticos se obtienen de:

- **Fuentes primarias*:*** obtenidos directamente sin intermediarios valiéndose de observaciones, encuestas, entrevistas y sondeos de opinión.
- **Fuentes secundarias*:*** obtenidos a través de intermediarios valiéndose de textos, revistas, documentos, publicaciones de prensa, y demás trabajos hechos por personas o entidades.

ESCALAS DE MEDICIÓN

En cuanto a las escalas de medición la estadística cuenta con las siguientes:

- **Nominal**

  Se utiliza principalmente en los datos cualitativos y nos permite manejar la información por su nombre, como en los casos de marcas de diferentes productos, enfermedades, preferencias, etc.

  <u>Por ejemplo:</u>

  - Sexo: las clases son masculino o femenino.
  - Especialidad: las diferentes especialidades (carreras) de la UCV.
  - Número de cédula de identidad personal.
  - Temperatura de una persona: sanguíneo, flemático, melancólico, colérico.

- **Ordinal**

  Se utiliza cuando necesitamos establecer orden entre las diferencias de la población y sus datos son cualitativos, por ejemplo, escalas de calidad (mala, regular, buena, muy buena), escalas de gusto (mu y sabrosa, sabrosa, agradable, desagradable, muy desagradable), etc.

Por ejemplo:

- Evaluaciones en un examen: 5, 4, 3 y 2.
- Grado de satisfacción de una necesidad: alto, medio, bajo
- Conocimiento de un idioma: excelente, bien, regular, mal

- **Intervalo**

  Espacio o distancia que hay de un tiempo a otro o de un lugar a otro. Se utiliza principalmente en datos cuantitativos y es una escala que no cuenta con un cero absoluto o con un instrumento estandarizado, por ejemplo, la temperatura se puede medir en grados centígrados, Fahrenheit o kelvin.

- **Razón**

  Básicamente utilizada en datos cuantitativos que pueden ser medidos con instrumentos estandarizados o con un cero absoluto como por ejemplo una distancia medida en kilómetros, un volumen medido en centímetros cúbicos, ventas medidas en bolívares, etc.

CENSO

Es una técnica de recolección de datos estadísticos que se realiza a toda la población

ENCUESTA

Es la técnica que nos permite recolectar datos estadísticos que se realiza a una muestra de la población.

Se clasifica en:

- ***Descriptiva.-*** Cuando registra datos referentes a las características de los elementos o individuos.
- ***Explicativa.-*** Cuando averigua las causas o razones que originan los fenómenos.
- ***Mixtas.-*** Cuando es descriptiva y explicativa.
- ***Por muestreo.-*** Cuando recolecta información de grupos representativos de la población.

Su estructura es:

- Nombre de la institución que auspicia la encuesta.
- Tema de la encuesta.
- *Objetivos* de la encuesta.
- Datos informativos: Lugar, fecha, y otros datos que se considere necesario según la naturaleza de la información estadística a encuestarse.
- *Instrucciones* para el encuestado para que sepa la forma de llenar la encuesta.
- Cuestionario o listado de preguntas (cerradas, abiertas, o ambas a la vez) sobre los diferentes aspectos motivo de estudio.
- Frase de agradecimiento al encuestado, como por ejemplo, ¡Gracias por su colaboración!

Las diferentes tipos de preguntas pueden ser:

- ***Abiertas.-*** Son aquellas en la cual el encuestado construye la respuesta de manera libre según su opinión y de la manera que él desea. Ejemplo:
  ¿Qué piensa usted sobre la política educativa del actual gobierno?

- ***Cerradas o dicotómicas.-*** Sólo pueden ser contestadas por un "sí" o por un "no". Ejemplo:
  ¿Está usted de acuerdo con la política educativa del actual gobierno?
  Si ( )   No ( )
  Como es obvio, la respuesta será forzosamente una de las alternativas planteadas: Las preguntas cerradas son fáciles de tabular y facilitan la cuantificación mediante la asignación de puntuaciones.

- ***Preguntas de elección múltiple o categorizada:*** Se trata en cierto modo de preguntas cerradas que, dentro de los extremos de una escala permiten una serie de alternativas de respuestas cuyos matices son fijados de antemano. Presentan dos formas: En abanico y de estimación.

  - ***Preguntas con respuesta en abanico:*** Estas preguntas permiten contestar señalando una o varias respuestas presentadas junto con la pregunta. Por ejemplo: Indique otras alternativas que considere importantes para mejorar la educación en nuestro país.

- ***Preguntas de Estimación:*** Son preguntas cuantitativas que introducen diversos grados de intensidad creciente o decreciente para un mismo ítem.

Ejemplos:

- ¿Cómo calificaría la política educativa del gobierno?
  Excelente ( ) Muy Buena ( ) Regular ( ) Deficiente ( )

- ¿Qué porcentaje está de acuerdo con la política educativa del gobierno?
  100% ( ) 75% ( ) 50% ( ) 25% ( ) 0% ( )

- ¿Le interesa conocer el modelo educativo vigente?
  Nada ( ) Poco ( ) Algo ( ) Mucho ( )

- ¿Piensa culminar sus estudios superiores?
  Sí ( ) Probablemente Sí ( ) No ( ) Aún no decido ( )

AUTOEVALUACIÓN

1-. Defina Estadística
2-. Mencione algunas aplicaciones de la estadística
3-. ¿Cuáles son los fines de la estadística?
4-. Diga los objetivos de la estadística
5-. Explique en secuencia las fases del método estadístico.
6-. ¿Qué es la Estadística Descriptiva o Deductiva?
7-. ¿Qué es la Estadística Inferencial o Inductiva?
8-. Defina Población
9-. ¿Cuándo una población es finita?
10-. ¿Cuándo una población es infinita?
11-. ¿Cuál es la diferencia entre población y muestra?
12-. Defina variable
13-. ¿Cuándo una Variable es cualitativa?
14-. ¿Cuándo una Variable es cuantitativa?
15-. ¿Qué es una Variable dicotómicas?
16-. ¿Qué es una Variable Politómicas?
17-. ¿Qué es una Variable Discreta?
18-. ¿Qué es una Variable Continua?
19-. ¿Qué es una Variable Ordinal?
20-. ¿Qué es una Variable Nominal?
21-. Defina Individuo
22-. ¿Qué es un parámetro?
23-. ¿Qué es un estadístico?
24-. Defina Datos estadísticos
25-. ¿Cómo se pueden clasificar los datos estadísticos?
26-. ¿En los datos estadísticos, cuándo la fuente es primaria o secundaria?
27-. Defina censo
28-. Defina encuesta
29-. ¿Cómo se clasifican las encuestas?
30-. ¿Cuándo una encuesta es explicativa?
31-. ¿Cuándo una pregunta es abierta?
32-. ¿Cuándo una pregunta es cerrada?
33-. ¿Cuándo una pregunta es de *elección múltiple*?
34-. ¿Cuándo una pregunta de *elección múltiple* permite *respuesta en abanico*?
35-. ¿Cuándo una pregunta de *elección múltiple* es *de Estimación*?

MUESTREO

Muestra

Es un subconjunto de la población. Ejemplo: Estudiantes de 2do lapso del IUDAG.

Sus principales características son:

- ***Representativa.-*** Se refiere a que todos y cada uno de los elementos de la población tengan la misma oportunidad de ser tomados en cuenta para formar dicha muestra.
- ***Adecuada y válida.-*** Se refiere a que la muestra debe ser obtenida de tal manera que permita establecer un mínimo de error posible respecto de la población.

CÁLCULO DE LA TALLA ÓPTIMA DE LA MUESTRA

Para que una muestra sea fiable, es necesario que su tamaño sea obtenido mediante procesos matemáticos que eliminen la incidencia del error.

Para calcular el tamaño de la muestra suele utilizarse la siguiente fórmula:

$$n = \frac{N\sigma^2 Z^2}{(N-1)\,e^2 + \sigma^2 Z^2}$$

Siendo:

- n = el tamaño de la muestra.
- N = tamaño de la población.
- σ = Desviación estándar de la población que, generalmente cuando no se tiene su valor, suele utilizarse un valor constante de 0,5.
- Z = Valor obtenido mediante niveles de confianza. Es un valor constante que, si no se tiene su valor, se lo toma en relación al 95% de confianza equivale a 1,96 (como más usual) o en relación al 99% de confianza equivale 2,58, valor que queda a criterio del encuestador.
- e = Límite aceptable de error muestral que, generalmente cuando no se tiene su valor, suele utilizarse un valor que varía entre el 1% (0,01) y 9% (0,09), valor que queda a criterio del encuestador.

Ejemplo ilustrativo: Calcular el tamaño de la muestra de una población de 1000 elementos.

Solución:

Se tiene N=1000, como no se tiene los demás valores se tomará:

$\sigma = 0,5$

$Z = 1,96$

$e = 0,05$.

Reemplazando valores en la fórmula se obtiene:

$$n = \frac{N\sigma^2Z^2}{(N-1)e^2 + \sigma^2Z^2} = \frac{1000 * 0,5^2 * 1,96^2}{(1000-1) * 0,05^2 * 1,96^2} = \frac{1000 * 0.25 * 3,8416}{(999) * 0.0025 + 0.25 * 3.8416}$$

$$n = \frac{960,4}{2,4975 + 0,9604} = \frac{960,4}{3,4579} = 277,74 \approx 278$$

MÉTODOS DE COLECTA DE DATOS

La recogida de datos es la primera etapa a pensar y decidir. La primera cuestión será el objetivo. La segunda es la del tipo de datos a recoger. Existen 3 tipos de recogida de datos:

- La recolección de datos periódica
- El muestreo o el censo
- La recogida mediante investigación científica

Sólo se presentaran aquí los métodos de muestreo que permiten recoger las informaciones necesarias para calcular las probabilidades sobre una población con un nivel de confianza definido.

El objetivo del muestro es el de recabar los datos necesarios para practicar la estadística inferencial.

Los distintos métodos son:

- El muestreo aleatorio simple
- El muestreo por conglomerados
- El muestreo por cuotas
- El muestreo estratificado
- El muestreo sistemático

Se pueden distinguir 2 tipos de métodos de muestreo: los no aleatorios y los aleatorios.

- El método no aleatorio más utilizado es el muestreo por cuotas: a partir de las especificidades de una población, se elabora la muestra respetando los niveles deseados eligiendo aleatoriamente los individuos con mismas características.
- Los métodos aleatorios recurren a los conceptos de probabilidades en la elección de las muestras. Solos estos métodos permiten estimar el nivel de confianza de los resultados de la población muestral.

Muestreo aleatorio simple

El muestreo aleatorio simple es la base de la teoría muestral.

Para obtener una muestra de este tipo, numeramos los individuos de la población de 1 a N, pues se extraen n individuos.

El objetivo es proporcionar una estimación insesgada de la media y de la varianza poblacional.

Muestreo por conglomerados

Para el muestreo por conglomerados es necesario dividir la población en sub-grupos.

Cada uno de esos sub grupos tiene que ser representativos de la población representada.

El muestreo por conglomerados consiste en poner aleatoriamente a los individuos escogidos en cada sub-grupo elegido y basar el estudio en estos individuos.

El ejemplo más común de este tipo de muestreo es un gran estudio de una ciudad que se divide asimismo en un estudio para cada barrio y entonces se efectúa un muestreo aleatorio simple.

Para obtener un muestreo por conglomerados manteniendo las propiedades estadísticas más precisas posibles, es necesario:

- Un número de conglomerados no consecuente.
- Tamaño de los conglomerados uniformes.
- Homogeneidad de los individuos dentro de los conglomerados.

Muestreo por cuotas

Puede que el método más utilizado actualmente, especialmente en los sondeos y encuestas utilizados por los medios de comunicación. Se trata de construir una muestra idéntica a la población a estudiar en términos de propiedades.

Por lo tanto se trata de un método no aleatorio.

El método de cuotas se basa en la distribución conocida de una población (edad, sexo, situación geográfica, categoría socio-profesional...).

Una vez determinada la dimensión del sondeo que se desea efectuar, basta con calcular el número de individuos por cada criterio elegido.

Sin embargo, este método (el menos costoso) tiene ciertas limitaciones que es necesario precisar y que nos permite conocer uno de los errores más comunes en las encuestas:

- Este método se basa en la hipótesis que la información que deseamos obtener está correlacionada con la población. Pero es solo una hipótesis de representatividad difícilmente demostrable.

- La elección de los individuos seleccionados por los investigadores en este método no permite calcular probabilidades de pertenencia a la muestra. Y entraña una dificultad de cálculo de errores y en consecuencia de precisión del análisis.

AUTOEVALUACIÓN

1-. Defina Muestra

2-. ¿A qué se refiere la característica representativa de la muestra?

3-. ¿Cuándo se dice que la característica de la muestra es adecuada y válida?

4-. ¿Cómo se calcula de la talla óptima de la muestra

5-. ¿Dentro de la fórmula para calcular el tamaño de la muestra, qué significa:

a. n
b. N
c. σ
d. Z
e. e

6-. ¿Cuándo no se tiene la Desviación estándar, qué constante se usa?

7-. ¿Cuándo no se tiene el nivel de confianza, qué valores se emplean?

8-. ¿Cuándo no se tiene el error muestral, qué valores se utilizan?

9-. Calcule la muestra para una población de 10.000 habitantes

10-. Explique los tres (3) tipos de recogida de datos

11-. Diga los distintos métodos de muestreo

12-. Cuál es el método no aleatorio más utilizado

13-. Qué es el métodos de colecta de datos

14-. Diga cómo se procede con cada uno de los siguientes métodos:

a. Muestreo aleatorio simple
b. Muestreo por conglomerados
c. Muestreo por cuotas

DISTRIBUCIÓN DE FRECUENCIA

Frecuencias

- ***Frecuencia Absoluta (f_i)***

  Número de veces que aparece un determinado valor en un estudio estadístico (veces que se repite), se representa por $\mathbf{f_i}$. La suma de las "$\mathbf{f_i}$" es igual al número total de datos, que se representa por **N**. => $\mathbf{\Sigma\, f_i = N}$

- ***Frecuencia Absoluta Acumulada (F_i)***

  En la primera clase es la misma frecuencia absoluta, a partir de la segunda se obtiene sumando la frecuencia absoluta de la primera clase con la de la segunda, este valor con la de la tercera, y así sucesivamente.

- ***Frecuencia Relativa (h_i)***

  Indica la proporción con que se repite un valor. Es el cociente entre la frecuencia absoluta de una determinada clase y el número total de datos. La suma de las frecuencias relativas es siempre 1. Se puede expresar en tantos por ciento y se representa por hi. => **hi** = **fi /N**

- ***Frecuencia Relativa Acumulada (H_i)***

  En la primera clase es la misma frecuencia relativa, a partir de la segunda se obtiene sumando la frecuencia relativa de la primera clase con la de la segunda, este valor con la de la tercera, y así sucesivamente.

Límite de la clase: cada clase está delimitado por el límite inferior (**Li**) de la clase y el límite superior (**Ls**) de la misma.

Amplitud de la clase: diferencia entre el límite superior e inferior de la clase.

Marca de clase: es el punto medio de cada intervalo y es el valor que representa a todo el intervalo para el cálculo de algunos parámetros, se determina con las semisuma de los límites de clase, es decir: $\dot{X} = (Li + Ls)/2$.

Reglas para construir las distribuciones de Frecuencias por intervalos

i. Se ordenan los datos de menor a mayor

ii. Se determina N (total de datos)

iii. Se consigue el dato mayor y dato menor => X_{max} y X_{min}

iv. Se determina la cantidad de clases o intervalos
- Mediante la fórmula Sturges => $k \approx 1 + 3{,}322 \, Log(N)$

v. Se calcula el intervalo total IT, AT, R (amplitud total, rango o recorrido)
- $IT = X_{max} - X_{min}$

  Se efectúa la resta, y si el resultado no es división exacta entre el número de intervalo, se busca el número entero mayor, inmediato a la diferencia, que sea divisible por el número de intervalos conseguido en el paso anterior.

vi. Se determina el intervalo de clase o amplitud => $Ic = IT \div k$

vii. Se arma la tabla partiendo del valor más bajo

<u>Conformación de las columnas</u>

a. Clases
b. Punto Medio de cada clase $(Li+Ls) \div 2$
c. Frecuencia Absoluta (fi)
d. Frecuencia absoluta acumulada (Fi)
e. Frecuencia relativa (hi) = $fi \div N$
f. Frecuencia relativa acumulada (Hi)

Nota:
Se forman los intervalos teniendo presente que el límite inferior de una clase **pertenece** al intervalo, pero el límite superior **no pertenece**, sino que se toma en cuenta en la siguiente clase. En la última clase sí se toma el límite superior.

Ejemplo ilustrativo:

Elaborar la tabla de frecuencia tomando en cuenta las edades comprendidas entre 1 y 40 años de 150 personas:

Suministran la serie de edades desordenadas (datos)

33	34	33	35	35	34	28	38	12	29
8	10	9	10	11	10	11	28	37	13
24	26	25	27	27	26	36	12	2	37
14	15	15	16	16	15	1	36	20	3
5	6	5	7	7	6	19	1	40	20
22	23	23	24	24	23	39	19	17	40
38	38	38	39	39	38	16	39	31	18
28	29	28	29	30	29	30	17	34	31
11	12	33	13	13	13	12	30	10	34
36	36	9	37	37	37	36	33	26	10
1	2	25	3	4	2	1	9	15	26
19	20	14	21	21	20	19	25	6	15
39	40	5	40	40	40	40	14	23	7
16	17	22	18	18	17	17	5	38	23
30	31	38	32	32	31	30	22	29	39

i. Se ordenan los datos de menor a mayor

1	1	1	1	2	2	2	3	3	4
5	5	5	5	6	6	6	7	7	7
8	9	9	9	10	10	10	10	10	11
11	11	12	12	12	12	13	13	13	13
14	14	14	15	15	15	15	15	16	16
16	16	17	17	17	17	17	18	18	18
19	19	19	19	20	20	20	20	21	21
22	22	22	23	23	23	23	23	24	24
24	25	25	25	26	26	26	26	27	27
28	28	28	28	29	29	29	29	29	30
30	30	30	30	31	31	31	31	32	32
33	33	33	33	34	34	34	34	35	35
36	36	36	36	36	37	37	37	37	37
38	38	38	38	38	38	38	39	39	39
39	39	39	40	40	40	40	40	40	40

ii. Se determina N (total de datos)

$N = 150$

iii. Se hallar el número de clases o intervalos de clases (K).

$K = 1 + 3.322(\log. N) = 1 + 3,322 (\log 150) = 8,2289 \approx 8$

iv. Se consigue el rango (R):

$$R = (X_{may} - X_{men}) = X_n - X_1 = 40 - 1 = 39 \approx 40$$

v. Se calcula el Intervalo o amplitud de la clase (Ic):

$$Ic = \frac{R}{K} = \frac{40}{8} = 5$$

vi. Se obtienes las frecuencias absolutas mediante la tabulación o conteo de los datos (homogenización de los datos)

X	**$\dot{X}$**	**f_i**	**$F_i=\Sigma f_i$**	**$h_i= f_i /N$**	**$H_i=\Sigma h_i$**
0 – 5	2,5	10	10	0,07	0,07
5 - 10	7,5	14	24	0,09	0,16
10 - 15	12,5	19	43	0,13	0,29
15 - 20	17,5	21	64	0,14	0,43
20 - 25	22,5	17	81	0,11	0,54
25-30	27,5	18	99	0,12	0,66
30-35	21,5	19	118	0,13	0,79
35-40	27,5	32	150	0,21	1
		N=150		**1**	

Ejercicios resueltos de frecuencia

1.- Durante el mes de julio, en una ciudad se han registrado las siguientes temperaturas máximas:

32, 31, 28, 29, 33, 32, 31, 30, 31, 31, 27, 28, 29, 30, 32, 31, 31, 30, 30, 29, 29, 30, 30, 31, 30, 31, 34, 33, 33, 29, 29.

Construir la tabla de frecuencias.

x_i	**f_i**	**F_i**	**n_i**	**N_i**
27	1	1	0.032	0.032
28	2	3	0.065	0.097
29	6	9	0.194	0.290
30	7	16	0.226	0.516
31	8	24	0.258	0.774
32	3	27	0.097	0.871
33	3	30	0.097	0.968
34	1	31	0.032	1
	31		**1**	

2.- Las puntuaciones obtenidas por un grupo de en una prueba han sido:

15, 20, 15, 18, 22, 13, 13, 16, 15, 19, 18, 15, 16, 20, 16, 15, 18, 16, 14, 13.

Construir la **tabla de distribución de frecuencias**.

x_i	Recuento	f_i	F_i	n_i	N_i
13	III	3	0.15	3	1
14	I	1	0.05	4	0.95
15	~~IIII~~	5	0.25	9	0.85
16	IIII	4	0.20	13	0.80
18	III	3	0.15	16	0.65
19	I	1	0.05	17	0.45
20	II	2	0.10	19	0.20
22	I	1	0.05	20	0.15
		20			

3.- El número de estrellas de los hoteles de una ciudad viene dado por la siguiente serie:

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

Construir la **tabla de distribución de frecuencias**.

x_i	Recuento	x_i	F_i	n_i	N_i
1	~~IIII~~ I	6	6	0.158	0.158
2	~~IIII~~ ~~IIII~~ II	12	18	0.316	0.474
3	~~IIII~~ ~~IIII~~ ~~IIII~~ I	16	34	0.421	0.895
4	IIII	4	38	0.105	1
		38		**1**	

4.- Las calificaciones de 50 alumnos en Matemáticas han sido las siguientes:

5, 2, 4, 9, 7, 4, 5, 6, 5, 7, 7, 5, 5, 2, 10, 5, 6, 5, 4, 5, 8, 8, 4, 0, 8, 4, 8, 6, 6, 3, 6, 7, 6, 6, 7, 6, 7, 3, 5, 6, 9, 6, 1, 4, 6, 3, 5, 5, 6, 7.

Construir la tabla de distribución de frecuencias.

x_i	f_i	F_i	n_i	N_i
0	1	1	0.02	0.02
1	1	2	0.02	0.04
2	2	4	0.04	0.08
3	3	7	0.06	0.14
4	6	13	0.12	0.26
5	11	24	0.22	0.48
6	12	36	0.24	0.72
7	7	43	0.14	0.86
8	4	47	0.08	0.94
9	2	49	0.04	0.98
10	1	50	0.02	1.00
	50		**1.00**	

5.- Los 40 alumnos de una clase han obtenido las siguientes puntuaciones, sobre 50, en un examen de Física.

3, 15, 24, 28, 33, 35, 38, 42, 23, 38, 36, 34, 29, 25, 17, 7, 34, 36, 39, 44, 31, 26, 20, 11, 13, 22, 27, 47, 39, 37, 34, 32, 35, 28, 38, 41, 48, 15, 32, 13.

Construir la **tabla de frecuencias**.

	x_i	f_i	F_i	n_i	N_i
[0, 5)	2.5	1	1	0.025	0.025
[5, 10)	7.5	1	2	0.025	0.050
[10, 15)	12.5	3	5	0.075	0.125
[15, 20)	17.5	3	8	0.075	0.200
[20, 25)	22.5	3	11	0.075	0.275
[25, 30)	27.5	6	17	0.150	0.425
[30, 35)	32.5	7	24	0.175	0.600
[35, 40)	37.5	10	34	0.250	0.850
[40, 45)	47.5	4	38	0.100	0.950
[45, 50)	47.5	2	40	0.050	1.000
		40		**1**	

6.- Los pesos de los 65 empleados de una fábrica vienen dados por la siguiente tabla:

Peso	**f_i**
[50, 60)	8
[60, 70)	10
[70, 80)	16
[80,90)	14
[90, 100)	10
[100, 110)	5
[110, 120)	2

Construir la **tabla de frecuencias**.

	x_i	**f_i**	**F_i**	**n_i**	**N_i**
[50, 60)	55	8	8	0.12	0.12
[60, 70)	65	10	18	0.15	0.27
[70, 80)	75	16	34	0.24	0.51
[80,90)	85	14	48	0.22	0.73
[90, 100)	95	10	58	0.15	0.88
[100, 110)	105	5	63	0.08	0.96
[110, 120)	115	2	65	0.03	0.99
		65			

7.- Un dentista observa el número de caries en cada uno de los 100 niños de cierto colegio. La información obtenida a parecer resumida en la siguiente tabla**:**

Nº de caries	**f_i**	**n_i**
0	25	0.25
1	20	0.2
2	x	z
3	15	0.15
4	y	0.05

Completar la tabla obteniendo los valores x, y, z.

La suma de las frecuencias relativas ha de ser igual a 1:

$$0.25 + 0.2 + z + 0.15 + 0.05 = 1$$

$$0.65 + z = 1 \quad \Rightarrow \quad z = 1 - 0{,}65 \quad \Rightarrow \quad z = 0.35$$

La frecuencia relativa de un dato es igual su frecuencia absoluta dividida entre 100, que es la suma de las frecuencias absolutas.

$$\frac{x}{100} = 0.35 \qquad x = 35 \qquad \frac{y}{100} = 0.05 \qquad y = 5$$

Nº de caries	f_i	n_i	$f_i \cdot n_i$
0	25	0.25	0
1	20	0.2	20
2	35	0.35	70
3	15	0.15	45
4	5	0.05	20
			155

8.- Completar los datos que faltan en la siguiente tabla estadística:

x_i	f_i	F_i	n_i
1	4		0.08
2	4		
3		16	0.16
4	7		0.14
5	5	28	
6		38	
7	7	45	
8			

4/N = 0,08 => N = 4/0,08 => **N** = 50
Primera fila: $\mathbf{F_1}$ = 4
Segunda fila: $\mathbf{F_2}$ = 4+4 = 8 $\mathbf{n_2}$ = 4/50 = 0,08
Tercera fila: $\mathbf{f_3}$ = n_3*50 = 0,16*50 = 8
Cuarta fila: $\mathbf{F_4}$ = 16 + 7 = 23
Quinta fila: $\mathbf{n_5}$ = 5/50 = 0,1
Sexta fila: $\mathbf{f_6}$ = F_6–F_5 = 38–28 = 10 $\mathbf{n_6}$ = 10/50 = 0,2
Séptima fila: $\mathbf{n_7}$ = 7/50 = 0,14
Octava fila: $\mathbf{f_8}$ = 50−45 = 5 $\mathbf{n_8}$ = 5/50 = 0,1

x_i	f_i	F_i	n_i	$x_i \cdot f_i$
1	4	4	0.08	4
2	4	8	0.08	8
3	8	16	0.16	24
4	7	23	0.14	28
5	5	28	0.1	25
6	10	38	0.2	60
7	7	45	0.14	49
8	5	50	0.1	40
	50			**238**

AUTOEVALUACIÓN

1-. Defina Frecuencia Absoluta

2-. Defina Frecuencia Relativa

3-. Defina Frecuencia Acumulada (fa)

4-. Defina Frecuencia Porcentual (f%)

5-. Defina Frecuencia Relativa Acumulada (fra)

6-. Defina Frecuencia Relativa Acumulada Porcentual (fra%)

7-. Determine las frecuencias en el cuadro siguientes:

10	8	5	8	9	8	1	10
6	7	8	9	4	8	10	8
6	5	3	8	10	5	4	9
8	10	6	7	3	7	4	6
8	10	7	8	5	9	38	5

8-. Defina rango y diga Cuál es la fórmula para calcularlo

9-. ¿Qué es y cómo se selecciona el número de intervalos de clase?

10-. ¿Qué es y cómo se calcula el ancho del intervalo?

11-. ¿Qué es y cómo se calcula la marca de clase?

MEDIDAS Y TENDENCIA

MEDIDAS DE TENDENCIA CENTRAL

PARÁMETROS ESTADÍSTICOS

Un parámetro estadístico es un número que se obtiene a partir de los datos de una distribución estadística.

Los parámetros estadísticos sirven para sintetizar la información dada por una tabla o por una gráfica.

Tipos de parámetros estadísticos

Hay tres tipos parámetros estadísticos:

- De centralización.
- De posición
- De dispersión.

MEDIDAS DE CENTRALIZACIÓN

Nos indican en torno a qué valor (centro) se distribuyen los datos. Las medidas de centralización son: Media Aritmética, Mediana y Moda

MEDIA ARITMÉTICA SIMPLE

Es la medida de tendencia central más utilizada por lo general se ubica hacia el centro de distribución estadística. Se obtiene al sumar todos los datos y dividir el resultado entre el número total de datos.

$\bar{X}$ es el símbolo de la media aritmética.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{N} \qquad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{N}$$

Ejemplo:

Los pesos de seis amigos son: 84, 91, 72, 68, 87 y 78 kg. Hallar el peso medio.

$$\bar{x} = \frac{84 + 91 + 72 + 68 + 87 + 78}{6} = 80\ Kg$$

MEDIA ARITMÉTICA PARA DATOS AGRUPADOS

Si los datos vienen agrupados en una tabla de frecuencias, la expresión de la media es:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \ldots + x_n f_n}{N} \qquad \bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{N}$$

Ejercicio de media aritmética

En un test realizado a un grupo de 42 personas se han obtenido las puntuaciones que muestra la tabla. Calcula la puntuación media.

	x_i	f_i	$x_i \cdot f_i$
[10, 20)	15	1	15
[20, 30)	25	8	200
[30,40)	35	10	350
[40, 50)	45	9	405
[50, 60	55	8	440
[60,70)	65	4	260
[70, 80)	75	2	150
		42	**1 820**

$$\bar{x} = \frac{1820}{42} = 43.33$$

<u>Ejercicios resueltos de la media aritmética</u>

1.- Considérense los siguientes datos: 3, 8, 4, 10, 6, 2. Se pide:

a.-Calcular su media.

$$\bar{x}_1 = \frac{33}{6} = 5.5$$

b.-Si los todos los datos anteriores los multiplicamos por 3, cúal será la nueva media.

$$\bar{x}_2 = 5.5 \cdot 3 = 16.5$$

2.- 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?

$$\bar{x} = \frac{5 \cdot 7.31 + 4.47 + 10.15}{7} = 7.31$$

3.- Calcular la media de una distribución estadística que viene dada por la siguiente tabla:

x_i	61	64	67	70	73
f_i	5	18	42	27	8

x_i	f_i	$x_i \cdot f_i$
61	5	305
64	18	1152
67	42	2814
71	27	1890
73	8	584
	100	**6745**

$$\bar{x} = \frac{6745}{100} = 67.45$$

4.- Hallar la media de la distribución estadística que viene dada por la siguiente tabla:

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

	x_i	f_i	$x_i \cdot f_i$
[10, 15)	12.5	3	37.5
[15, 20)	17.5	5	87.5
[20, 25)	22.5	7	157.5
[25, 30)	27.5	4	110
[30, 35)	32.5	2	65
		21	**457.5**

$$\bar{x} = \frac{457.5}{21} = 21.79$$

LA MEDIANA.-

La mediana, es el valor que divide una distribución de datos ordenados en dos partes iguales, es decir, el 50% de los datos se ubican sobre la mediana o hacia los puntajes altos y el 50% restante hacia los puntajes bajos. Es el valor que ocupa el lugar central de todos los datos cuando éstos están ordenados de menor a mayor.

La mediana se representa por $\mathbf{M_e}$.

La mediana se puede hallar sólo para variables cuantitativas.

Cálculo de la mediana

1. Ordenamos los datos de menor a mayor.

2. Si la serie tiene un número impar de medidas la mediana es la puntuación central de la misma.

   2, 3, 4, 4, 5, 5, 5, 6, 6 Me = 5

3. Si la serie tiene un número par de puntuaciones la mediana es la media entre las dos puntuaciones centrales.

   7, 8, 9, 10, 11, 12 Me = 9.5

Cálculo de la mediana para datos agrupados

La mediana se encuentra en el intervalo donde la frecuencia acumulada llega hasta la mitad de la suma de las frecuencias absolutas.

Es decir tenemos que buscar el intervalo en el que se encuentre N/2

$$Me = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot a_i$$

Descripción:

$\mathbf{L_i}$ es el límite inferior de la clase donde se encuentra la mediana.

$\frac{N}{2}$ es la semisuma de las frecuencias absolutas.

$\mathbf{F_{i-1}}$ es la frecuencia acumulada anterior a la clase mediana.

$\mathbf{a_i}$ es la amplitud de la clase.

La mediana es independiente de las amplitudes de los intervalos.

Ejemplo:

Calcular la mediana de una distribución estadística que viene dada por la siguiente tabla:

	f_i	F_i
[60, 63)	5	5
[63, 66)	18	23
[66, 69)	42	65
[69, 72)	27	92
[72, 75)	8	100
	100	

100/2 = 50

Clase de la mediana: [66, 69)

$$Me = 66 + \frac{50 - 23}{42} \cdot 3 = 67.93$$

Ejercicios resueltos de la mediana

- Hallar la **mediana** de la siguientes series de números:

  a.- 3, 5, 2, 6, 5, 9, 5, 2, 8. => 2, 2, 3, 5, 5, 5, 6, 8, 9. => Me = 5

  b.- 3, 5, 2, 6, 5, 9, 5, 2, 8, 6. => 2, 2, 3, 5, 5, 5, 6, 6, 8, 9. => $Me = \frac{5+5}{2} = 5$

  c.- 10, 13, 4, 7, 8, 11 10, 16, 18, 12, 3, 6, 9, 9, 4, 13, 20, 7, 5, 10, 17, 10, 16, 14, 8, 18

  3, 4, 4, 5, 6, 7, 7, 8, 8, 9, 9, 10, 10, 10, 10, 11, 12, 13, 13, 14, 16, 16, 17, 18, 18, 20

$$Me = \frac{10+10}{2} = 10$$

- Tabular y calcular mediana de la siguiente serie de números: 5, 3, 6, 5, 4, 5, 2, 8, 6, 5, 4, 8, 3, 4, 5, 4, 8, 2, 5, 4.

x_i	f_i	F_i
2	2	2
3	2	4
4	5	9
5	6	15
6	2	17
8	3	20
	20	

20/2 = 10 Me = 5

- Hallar la mediana de la distribución estadística que viene dada por la siguiente tabla:

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

	f_i	F_i
[10, 15)	3	3
[15, 20)	5	8
[20, 25)	7	15
[25, 30)	4	19
[30, 35)	2	21
	21	

$$\frac{21}{2} = 10.5 \qquad Me = 20 + \frac{10.5 - 8}{7} \cdot 5 = 21.786$$

MODA

La moda de un conjunto de datos es el valor que aparece con mayor frecuencia.

Se representa por **Mo**.

Se puede hallar la moda para variables cualitativas y cuantitativas.

Para un conjunto de datos unimodales existe la siguiente relación empírica:

Media aritmética – moda = 3 (media aritmética – mediana)

Hallar la **moda** de la distribución:

2, 3, 3, 4, 4, 4, 5, 5 $M_o = 4$

Si en un grupo hay dos o varias puntuaciones con la misma frecuencia y esa frecuencia es la máxima, la distribución es bimodal o multimodal, es decir, tiene varias modas.

1, 1, 1, 4, 4, 5, 5, 5, 7, 8, 9, 9, 9 M_o= 1, 5, 9

Cuando todas las puntuaciones de un grupo tienen la misma frecuencia, no hay moda.

2, 2, 3, 3, 6, 6, 9, 9

Si dos puntuaciones adyacentes tienen la frecuencia máxima, la moda es el promedio de las dos puntuaciones adyacentes.

0, 1, 3, 3, 5, 5, 7, 8 Mo = 4

Cálculo de la moda para datos agrupados

Todos los intervalos tienen la misma amplitud.

$$Mo = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot a_i$$

Descripción:

L_i es el límite inferior de la clase modal.

f_i es la frecuencia absoluta de la clase modal.

f_{i-1} es la frecuencia absoluta inmediatamente inferior a la clase modal.

f_{i+1} es la frecuencia absoluta inmediatamente posterior a la clase modal.

a_i es la amplitud de la clase.

También se utiliza otra fórmula de la moda que da un valor aproximado de ésta:

$$Mo = L_i + \frac{f_{i+1}}{f_{i-1} + f_{i+1}} \cdot a_i$$

Ejemplo:

Calcular la moda de una distribución estadística que viene dada por la siguiente tabla:

	f_i
[60, 63)	5
[63, 66)	18
[66, 69)	42
[69, 72)	27
[72, 75)	8
	100

$$Mo = 66 + \frac{(42-18)}{(42-18)+(42-27)} \cdot 3 = 67.846$$

$$Mo = 66 + \frac{27}{18+27} \cdot 3 = 67.8$$

Los intervalos tienen amplitudes distintas.

En primer lugar tenemos que hallar las alturas.

$$h_i = \frac{f_i}{a_i}$$

La clase modal es la que tiene mayor altura.

$$Mo = L_i + \frac{h_i - h_{i-1}}{(h_i - h_{i-1}) + (h_i - h_{i+1})} \cdot a_i$$

La fórmula de la moda aproximada cuando existen distintas amplitudes es:

$$Mo = L_i + \frac{h_{i+1}}{h_{i-1} + h_{i+1}} \cdot a_i$$

Ejemplo:

En la siguiente tabla se muestra las calificaciones (suspenso, aprobado, notable y sobresaliente) obtenidas por un grupo de 50 alumnos. Calcular la moda.

	f_i	h_i
[0, 5)	15	3
[5, 7)	20	10
[7, 9)	12	6
[9, 10)	3	3
	50	

$$Mo = 5 + \frac{10-3}{(10-3)+(10-6)} \cdot 2 = 6.27 \qquad Mo = 5 + \frac{6}{3+6} \cdot 2 = 6.33$$

<u>**Ejercicios resueltos de la moda**</u>

- Calcular la moda de la siguiente serie de números: 5, 3, 6, 5, 4, 5, 2, 8, 6, 5, 4, 8, 3, 4, 5, 4, 8, 2, 5, 4. => Mo = 5

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

Al calcular la moda nos queda Mo = 12

- Calcular la moda de una distribución estadística que viene dada por la siguiente tabla:

	f_i
[60, 63)	5
[63, 66)	18
[66, 69)	42
[69, 72)	27
[72, 75)	8
	100

$$Mo = 66 + \frac{(42-18)}{(42-18)+(42-27)} \cdot 3 = 67.846$$

- Calcular la moda de una distribución estadística viene dada por la siguiente tabla:

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

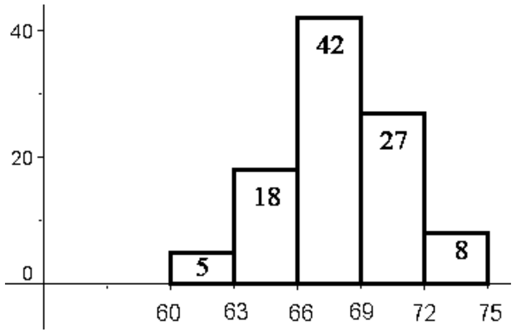
$$Mo = 20 + \frac{2}{2+3} \cdot 5 = 22$$

- Calcular la moda de la distribución estadística:

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

$$Mo = 15 + \frac{1}{1+6} \cdot 5 = 15.71$$

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



Al calcular la moda nos da => $Mo = 66 + \frac{24}{24+15} \cdot 3 = 67.85$

- En la siguiente tabla se muestra las calificaciones (suspenso, aprobado, notable y sobresaliente) obtenidas por un grupo de 50 alumnos. Calcular la moda.

	f_i	h_i
[0, 5)	15	3
[5, 7)	20	10
[7, 9)	12	6
[9, 10)	3	3
	50	

$$Mo = 5 + \frac{10-3}{(10-3)+(10-6)} \cdot 2 = 6.27$$

MEDIDAS DE POSICIÓN

Las medidas de posición dividen un conjunto de datos en grupos con el mismo número de individuos.

Para calcular las medidas de posición es necesario que los datos estén ordenados de menor a mayor.

Las medidas de posición son: Cuartiles, Deciles, Percentiles

Para calcular las medidas de posición es necesario que los datos estén ordenados de menor a mayor.

CUARTILES.-

Son cada uno de los 3 valores Q_1, Q_2, Q_3 que dividen a la distribución de los datos en 4 partes iguales.

Los cuartiles son un caso particular de los percentiles. Hay 3 cuartiles:

Primer cuartil: $Q_1 = P_{25}$

Segundo cuartil: $Q_2 = D_5 = P_{50} =$ Mediana

Tercer cuartil: $Q_3 = P_{75}$

Q_2 coincide con la **mediana**.

Cálculo de los cuartiles

1. Ordenamos los datos de menor a mayor.
2. Buscamos el lugar que ocupa cada cuartil mediante la expresión.

- **Número impar de datos 2, 5, 3, 6, 7, 4, 9**

$$\begin{array}{ccccccc} 2, & 3, & 4, & 5, & 6, & 7, & 9 \\ & \downarrow & & \downarrow & & \downarrow & \\ & Q_1 & & Q_2 & & Q_3 & \end{array} \qquad \frac{k \cdot N}{4},\ k = 1, 2, 3$$

- Número par de datos **2, 5, 3, 4, 6, 7, 1, 9**

$$\begin{array}{ccccc} 1, & \underline{2,\ 3}, & \underline{4,\ 5}, & \underline{6,\ 7}, & 9 \\ & 2.5 & 4.5 & 6.5 & \\ & \downarrow & \downarrow & \downarrow & \\ & Q_1 & Q_2 & Q_3 & \end{array}$$

- **Cálculo de los cuartiles para datos agrupados**

En primer lugar buscamos la **clase** donde se encuentra $\frac{k \cdot N}{4},\ k = 1, 2, 3$, en la **tabla de las frecuencias acumuladas**.

$$Q_k = L_i + \frac{\frac{k \cdot N}{4} - F_{i-1}}{f_i} \cdot a_i \qquad k = 1, 2, 3$$

L_i es el límite inferior de la clase donde se encuentra el cuartil.

N es la suma de las frecuencias absolutas.

F_{i-1} es la frecuencia acumulada anterior a la clase del cuartil.

a_i es la amplitud de la clase.

➢ Ejercicio de cuartiles

Calcular los cuartiles de la distribución de la tabla:

$$Q_k = L_i + \frac{\frac{k \cdot N}{4} - F_{i-1}}{f_i} \cdot a_i \qquad k = 1,2,3$$

	f_i	F_i
[50, 60)	8	8
[60, 70)	10	18
[70, 80)	16	34
[80, 90)	14	48
[90, 100)	10	58
[100, 110)	5	63
[110, 120)	2	65
	65	

➢ Cálculo del primer cuartil

$$\frac{65 \cdot 1}{4} = 16.25$$

$$Q_1 = 60 + \frac{16.25 - 8}{10} \cdot 10 = 68.25$$

➢ Cálculo del segundo cuartil

$$\frac{65 \cdot 2}{4} = 32.5$$

$$Q_2 = 70 + \frac{32.5 - 18}{16} \cdot 10 = 79.0625$$

➢ Cálculo del tercer cuartil

$$\frac{65 \cdot 3}{4} = 48.75$$

$$Q_3 = 90 + \frac{48.75 - 48}{10} \cdot 10 = 90.75$$

DECILES

Son cada uno de los 9 valores D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9 que dividen a la distribución de los datos en 10 partes iguales.

Los deciles dan los valores correspondientes al 10%, al 20%... y al 90% de los datos.

Primer decil es igual al décimo percentil ($D_1 = P_{10}$)

Segundo decil es igual al veinteavo percentil ($D_2 = P_{20}$), y así sucesivamente. **D_5** coincide con la mediana.

- **Cálculo de los deciles**

En primer lugar buscamos la clase donde se encuentra, en la tabla de las frecuencias acumuladas.

$$D_k = L_i + \frac{\frac{k \cdot N}{10} - F_{i-1}}{f} \cdot a_i \qquad k = 1, 2, ..9 \qquad \frac{k \cdot N}{10}, \; k = 1, 2, ..9$$

L_i es el límite inferior de la clase donde se encuentra el decil.

N es la suma de las frecuencias absolutas.

F_{i-1} es la frecuencia acumulada anterior a la clase el decil..

a_i es la amplitud de la clase.

- **Ejercicio de deciles**

Calcular los deciles de la distribución de la tabla:

	f_i	F_i
[50, 60)	8	8
[60, 70)	10	18
[70, 80)	16	34
[80, 90)	14	48
[90, 100)	10	58
[100, 110)	5	63
[110, 120)	2	65
	65	

➢ **Cálculo del primer decil**

$$\frac{65\cdot 1}{10}=6.5$$

$$D_1=50+\frac{6.5-0}{8}\cdot 10=58.12$$

➢ **Cálculo del segundo decil**

$$\frac{65\cdot 2}{10}=13$$

$$D_2=60+\frac{13-8}{10}\cdot 10=65$$

➢ **Cálculo del tercer decil**

$$\frac{65\cdot 3}{10}=19.5$$

$$D_3=70+\frac{19.5-18}{16}\cdot 10=70.94$$

➢ **Cálculo del cuarto decil**

$$\frac{65\cdot 4}{10}=26$$

$$D_4=70+\frac{26-18}{16}\cdot 10=75$$

➢ **Cálculo del quinto decil**

$$\frac{65\cdot 5}{10}=32.5$$

$$D_5=70+\frac{32.5-18}{16}\cdot 10=79.06$$

➢ **Cálculo del sexto decil**

$$\frac{65\cdot 6}{10}=39$$

$$D_6=80+\frac{39-34}{14}\cdot 10=83.57$$

➢ **Cálculo del séptimo decil**

$$\frac{65\cdot 7}{10}=45.5$$

$$D_7=80+\frac{45.5-34}{14}\cdot 10=88.21$$

➢ **Cálculo del octavo decil**

$$\frac{65\cdot 8}{10}=52$$

$$D_8=90+\frac{52-48}{10}\cdot 10=94$$

➢ **Cálculo del noveno decil**

$$\frac{65\cdot 9}{10}=58.5 \qquad D_9=100+\frac{58.5-58}{5}\cdot 10=101$$

PERCENTILES O CENTILES

Son cada uno de los 99 valores P_1, P_2, P_3,……..P_{99} que dividen a la distribución de los datos en 100 partes iguales.

Los percentiles dan los valores correspondientes al 1%, al 2%... y al 99% de los datos.

P_{50} coincide con la **mediana**.
P_{50} coincide con **D_5.**

➢ **Cálculo de los percentiles**

En primer lugar buscamos la clase donde se encuentra, en la tabla de las frecuencias acumuladas.

$$P_k=L_i+\frac{\frac{k\cdot N}{100}-F_{i-1}}{f_i}\cdot a_i \qquad \frac{k\cdot N}{100},\quad k=1,2,..99$$

L_i es el límite inferior de la clase donde se encuentra el percentil.
N es la suma de las frecuencias absolutas.
F_{i-1} es la frecuencia acumulada anterior a la clase del percentil.
a_i es la amplitud de la clase.

- **Ejercicio de percentiles**

Calcular el percentil 35 y 60 de la distribución de la tabla:

	f_i	F_i
[50, 60)	8	8
[60, 70)	10	18
[70, 80)	16	34
[80, 90)	14	48
[90, 100)	10	58
[100, 110)	5	63
[110, 120)	2	65
	65	

- **Percentil 35**

$$\frac{65 \cdot 35}{100} = 22.75 \qquad P_{35} = 70 + \frac{22.75 - 18}{16} \cdot 10 = 72.97$$

- **Percentil 60**

$$\frac{65 \cdot 60}{100} = 39 \qquad P_{60} = 80 + \frac{39 - 34}{14} \cdot 10 = 83.57$$

MEDIDAS DE DISPERSIÓN

Las medidas de dispersión nos informan sobre cuanto se alejan del centro los valores de la distribución.

Las medidas de dispersión son:

- **Rango o recorrido**

El rango es la diferencia entre el mayor y el menor de los datos de una distribución estadística.

- **Desviación media**

La desviación **media** es la media aritmética de los valores absolutos de las desviaciones respecto a la media.

- **Varianza**

La varianza es la media aritmética del cuadrado de las desviaciones respecto a la media.

- **Desviación típica**

La desviación típica es la raíz cuadrada de la varianza.

RANGO, AMPLITUD TOTAL O RECORRIDO

Dada una serie de valores x_1, x_2,… x_n, su recorrido es la diferencia aritmética entre el máximo y el mínimo de estos valores.

$$Re = Xmáx – Xmín$$

Es la medida de dispersión más sencilla y también, por tanto, la que proporciona menos información. Además, esta información puede ser errónea, pues el hecho de que no influyan más de dos valores del total de la serie puede provocar una deformación de la realidad

DESVIACIÓN MEDIA

➢ **Desviación respecto a la media**

La desviación respecto a la media es la diferencia en valor absoluto entre cada valor de la variable estadística y la media aritmética. **D_i = |x - x|**

➢ **Desviación media**

La desviación media es la media aritmética de los valores absolutos de las desviaciones respecto a la media.

La **desviación media** se representa por $D_{\bar{x}}$

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \ldots + |x_n - \bar{x}|}{N} \qquad D_{\bar{x}} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{N}$$

Ejemplo:

Calcular la desviación media de la distribución: 9, 3, 8, 8, 9, 8, 9, 18

$$\bar{x} = \frac{9+3+8+8+9+8+9+18}{8} = 9 \qquad D_{\bar{x}} = \frac{|9-9|+|3-9|+|8-9|+|8-9|+|9-9|+|8-9|+|9-9|+|18-9|}{8} = 2.25$$

➢ **Desviación media para datos agrupados**

Si los datos vienen agrupados en una tabla de frecuencias, la expresión de la desviación media es:

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}| f_1 + |x_2 - \bar{x}| f_2 + \ldots + |x_n - \bar{x}| f_n}{N} \qquad D_{\bar{x}} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}| f_i}{N}$$

Ejemplo:

Calcular la desviación media de la distribución:

	x_i	f_i	$x_i \cdot f_i$	|x - x|	|x - x| · f_i
[10, 15)	12.5	3	37.5	9.286	27.858
[15, 20)	17.5	5	87.5	4.286	21.43
[20, 25)	22.5	7	157.5	0.714	4.998
[25, 30)	27.5	4	110	5.714	22.856
[30, 35)	32.5	2	65	10.714	21.428
		21	457.5		98.57

$$\bar{x} = \frac{457.5}{21} = 21.786 \qquad D_{\bar{x}} = \frac{98.57}{21} = 4.69$$

VARIANZA

La varianza es la media aritmética del cuadrado de las desviaciones respecto a la media de una distribución estadística.

La varianza se representa por σ^2.

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{N} \qquad \sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{N}$$

➢ **Varianza para datos agrupados**

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \ldots + (x_n - \bar{x})^2 f_n}{N} \qquad \sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 f_i}{N}$$

Para simplificar el cálculo de la varianza vamos o utilizar las siguientes expresiones que son equivalentes a las anteriores.

$$\sigma^2 = \frac{x_1^2 + x_2^2 + \ldots + x_n^2}{N} - \bar{x}^2 \qquad \sigma^2 = \sum_{i=1}^{n}\frac{x_i^2}{N} - \bar{x}^2$$

➢ **Ejercicios de varianza**

Ejercicio 1:

Calcular la varianza de la distribución:

9, 3, 8, 8, 9, 8, 9, 18

$$\bar{x} = \frac{9+3+8+8+9+8+9+18}{8} = 9$$

$$\sigma^2 = \frac{(9-9)^2 + (3-9)^2 + (8-9)^2 + (8-9)^2 + (9-9)^2 + (8-9)^2 + (9-9)^2 + (18-9)^2}{8} = 15$$

Ejercicio 2:

Calcular la varianza de la distribución de la tabla:

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

	x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[10, 20)	15	1	15	225
[20, 30)	25	8	200	5000
[30,40)	35	10	350	12 250
[40, 50)	45	9	405	18 225
[50, 60	55	8	440	24 200
[60,70)	65	4	260	16 900
[70, 80)	75	2	150	11 250
		42	**1 820**	**88 050**

$$\bar{x} = \frac{1820}{42} = 43.33 \qquad \sigma^2 = \frac{88050}{42} - 43.33^2 = 218.94$$

DESVIACIÓN TÍPICA

La desviación típica es la raíz cuadrada de la varianza.

Es decir, la raíz cuadrada de la media de los cuadrados de las puntuaciones de desviación.

La desviación típica se representa por **σ**.

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{N}} \qquad \sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{N}}$$

➢ **Desviación típica para datos agrupados**

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \ldots + (x_n - \bar{x})^2 f_n}{N}} \qquad \sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 f_i}{N}}$$

Para simplificar el cálculo vamos o utilizar las siguientes expresiones que son equivalentes a las anteriores.

$$\sigma = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{N} - \bar{x}^2} \qquad \sigma = \sqrt{\sum_{i=1}^{n} \frac{x_i^2}{N} - \bar{x}^2}$$

$$\sigma = \sqrt{\frac{x_1^2 f + x_2^2 f_2 + \ldots + x_n^2 f_n}{N} - \bar{x}^2}$$

$$\sigma = \sqrt{\sum_{i=1}^{n} \frac{x_i^2 f_i}{N} - \bar{x}^2}$$

➢ Ejercicios de desviación típica

Ejercicio 1:

Calcular la desviación típica de la distribución:

9, 3, 8, 8, 9, 8, 9, 18

$$\bar{x} = \frac{9+3+8+8+9+8+9+18}{8} = 9$$

$$\sigma = \sqrt{\frac{(9-9)^2+(3-9)^2+(8-9)^2+(8-9)^2+(9-9)^2+(8-9)^2+(9-9)^2+(18-9)^2}{8}} = 3.87$$

Ejercicio 2:

Calcular la desviación típica de la distribución de la tabla:

	x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[10, 20)	15	1	15	225
[20, 30)	25	8	200	5000
[30,40)	35	10	350	12 250
[40, 50)	45	9	405	18 225
[50, 60)	55	8	440	24 200
[60,70)	65	4	260	16 900
[70, 80)	75	2	150	11 250
		42	**1 820**	**88 050**

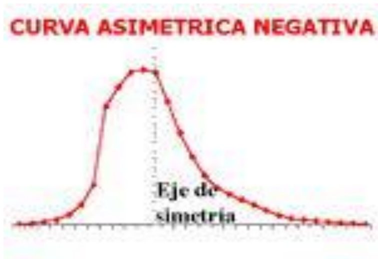
$$\bar{x} = \frac{1820}{42} = 43.33$$

$$\sigma = \sqrt{\frac{88050}{42} - 43.33^2} = 14.797$$

MEDIDAS DE FORMA

COEFICIENTE DE ASIMETRÍA

Asimetría se refiere a si la curva que forman los valores de la serie presenta la misma forma a izquierda y derecha de un valor central (media aritemética)



Para medir el nivel de asimetría se utiliza el llamado Coeficiente de Asimetría de Fisher, que viene definido:

$$g_1 = \frac{(1/n) * \Sigma (x_i - xm)^3 * n_i}{((1/n) * \Sigma (x_i - xm)^2 * n_i))^{3/2}}$$

Los resultados pueden ser los siguientes:

g1 = 0 (distribución simétrica; existe la misma concentración de valores a la derecha y a la izquierda de la media)

g1 > 0 (distribución asimétrica positiva; existe mayor concentración de valores a la derecha de la media que a su izquierda)

g1 < 0 (distribución asimétrica negativa; existe mayor concentración de valores a la izquierda de la media que a su derecha)

Ejemplo: Vamos a calcular el Coeficiente de Asimetría de Fisher de la serie de datos referidos a la estatura de un grupo de alumnos (lección 2ª):

Variable	Frecuencias absolutas		Frecuencias relativas	
(Valor)	Simple	Acumulada	Simple	Acumulada
1,20	1	1	3,3%	3,3%
1,21	4	5	13,3%	16,6%
1,22	4	9	13,3%	30,0%
1,23	2	11	6,6%	36,6%
1,24	1	12	3,3%	40,0%
1,25	2	14	6,6%	46,6%
1,26	3	17	10,0%	56,6%
1,27	3	20	10,0%	66,6%
1,28	4	24	13,3%	80,0%
1,29	3	27	10,0%	90,0%
1,30	3	30	10,0%	100,0%

Recordemos que la media de esta muestra es 1,253

□((xi - x)^3)*ni	□((xi - x)^2)*ni
0,000110	0,030467

Luego:

$$g1 = \frac{(1/30) * 0,000110}{(1/30) * (0,030467)^{(3/2)}} = -0,1586$$

Por lo tanto el Coeficiente de Fisher de Simetría de esta muestra es -0,1586, lo que quiere decir que presenta una distribución asimétrica negativa (se concentran más valores a la izquierda de la media que a su derecha).

COEFICIENTE DE CURTOSIS

El **Coeficiente de Curtosis** analiza el grado de concentración que presentan los valores alrededor de la zona central de la distribución.

Se definen 3 tipos de distribuciones según su grado de curtosis:

Distribución mesocúrtica: presenta un grado de concentración medio alrededor de los valores centrales de la variable (el mismo que presenta una distribución normal).

Distribución leptocúrtica: presenta un elevado grado de concentración alrededor de los valores centrales de la variable.

Distribución platicúrtica: presenta un reducido grado de concentración alrededor de los valores centrales de la variable.

El **Coeficiente de Curtosis** viene definido por la siguiente fórmula:

$$g_2 = \frac{(1/n) * \Sigma (x_i - xm)^4 * n_i}{((1/n) * \Sigma (x_i - xm)^2 * n_i))^2} - 3$$

Los resultados pueden ser los siguientes:

g2 = 0 (**distribución mesocúrtica)**.

g2 > 0 (distribución leptocúrtica).

g2 < 0 (distribución platicúrtica).

Ejemplo: Vamos a calcular el Coeficiente de Curtosis de la serie de datos referidos a la estatura de un grupo de alumnos (lección 2ª):

Variable	**Frecuencias absolutas**		**Frecuencias relativas**	
(Valor)	**Simple**	**Acumulada**	**Simple**	**Acumulada**
1,20	1	1	3,3%	3,3%
1,21	4	5	13,3%	16,6%
1,22	4	9	13,3%	30,0%
1,23	2	11	6,6%	36,6%
1,24	1	12	3,3%	40,0%
1,25	2	14	6,6%	46,6%
1,26	3	17	10,0%	56,6%
1,27	3	20	10,0%	66,6%
1,28	4	24	13,3%	80,0%
1,29	3	27	10,0%	90,0%
1,30	3	30	10,0%	100,0%

Recordemos que la media de esta muestra es 1,253

□((xi - xm)^4)*ni	**□((xi - xm)^2)*ni**
0,00004967	0,03046667

Luego:

$$g2 = \frac{(1/30) * 0,00004967}{((1/30) * (0,03046667))^2} - 3 = -1,39$$

Por lo tanto, el **Coeficiente de Curtosis** de esta muestra es -1,39, lo que quiere decir que se trata de una distribución platicúrtica, es decir, con una reducida concentración alrededor de los valores centrales de la distribución.-

MODELOS DE REGRESIÓN

CORRELACIÓN Y REGRESIÓN

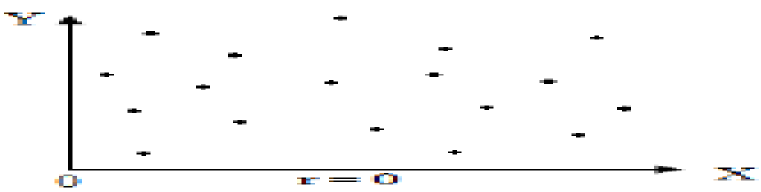
Cuando se estudian en forma conjunta dos características (variables estadísticas) de una población o muestra, se dice que estamos analizando una variable estadística bidimensional. La correlación es el grado de relación que existe entre ambas características, y la regresión es la forma de expresar matemáticamente dicha relación.

ANÁLISIS DE CORRELACIÓN.-

Dado dos variables, la correlación permite hacer estimaciones del valor de una de ellas conociendo el valor de la otra variable.

COEFICIENTES DE CORRELACIÓN.-

Los coeficientes de correlación son medidas que indican la situación relativa de los mismos sucesos respecto a las dos variables (Correlación de Karl Pearson y por Rangos de Spearman), es decir, son la expresión numérica que nos indica el grado de relación existente entre las 2 variables y en qué medida se relacionan. Son números que varían entre los límites +1 y -1. Su magnitud indica el grado de asociación entre las variables; el valor r = 0 indica que no existe relación entre las variables; los valores ± 1 son indicadores de una correlación perfecta positiva (al crecer o decrecer X, crece o decrece Y) o negativa (Al crecer o decrecer X, decrece o crece Y).



No hay correlación

Correlación Positiva

Correlación Negativa

Coeficiente de determinación

Revela qué porcentaje del cambio en Y se explica por un cambio en X. Se calcula elevando al cuadrado el coeficiente de correlación.

ANÁLISIS DE REGRESIÓN.-

Los primeros y más importantes estudios al respecto se deben a los científicos Francis Galton (1822-1911) y Karl Pearson (1857-1936). Fue Galton quien utilizó por primera vez el término regresión para indicar que, aunque influida por la estatura de sus padres, la estatura de los hijos "regresaba" a la media general.

La regresión examina la relación entre dos variables, pero restringiendo una de ellas con el objeto de estudiar las variaciones de una variable cuando la otra permanece constante. En otras palabras, la regresión es un método que se emplea para predecir el valor de una variable en función de valores dados a la otra variable. En estadística la palabra predecir no se utiliza en el sentido empleado por los astrólogos, futurólogos y mentalistas, sino más bien en un sentido lógico como es el de utilizar el conocimiento del comportamiento de una variable para obtener información sobre otra variable. Por ejemplo, puede predecirse el resultado que obtendrá un estudiante en su examen final, basados en el conocimiento de las calificaciones promedio de sus exámenes parciales, o predecir la preferencia de los estudiantes por profesiones científicas, conociendo los promedios de sus calificaciones en los estudios escolares.

En todos los casos de regresión existe una dependencia funcional entre las variables. En el caso de dos variables, siendo una de ellas (X) variable independiente y la otra (Y) la dependiente, se habla de egresión de Y sobre X;

Por ejemplo, los ingenieros forestales utilizan la regresión de la altura de los árboles sobre su diámetro, lo cual significa que midiendo el diámetro (variable independiente) y reemplazando su valor en una relación definida según la clase de árbol se obtiene la altura, y aun sin necesidad de cálculos aprecian la altura utilizando gráficas de la función de dependencia, altura = función del diámetro. *Suele emplearse el principio de los mínimos cuadrados*

LA RECTA DE LOS MÍNIMOS CUADRADOS.-

Se llama línea de mejor ajuste y se define como la línea que hace mínima la suma de los cuadrados de las desviaciones respecto a ella de todos los puntos que corresponden a la información recogida.

El punto de intersección entre las $Y = a_0 + a_1X$ rectas con $X = b_0 + b_1Y$ se simboliza $(\bar{X}, \bar{Y})$ y se llama **centroide** o centro de gravedad

LA PARÁBOLA DE LOS MÍNIMOS CUADRADOS.-

La parábola de mínimos cuadrados que aproxima el conjunto de puntos $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \dots (Y_N, Y_N)$ tiene ecuación dada por $Y = a_0 + a_1X + a_2X^2$, donde las constantes a_0, a_1 y a_2, se determinan al resolver simultáneamente el sistema de ecuaciones que se forma al multiplicar la ecuación por sucesivamente, y sumando después.

$$\begin{cases} \Sigma Y = a_0N + a_1\Sigma X + a_2\Sigma X^2 \\ \Sigma XY = a_0\Sigma X + a_1\Sigma X^2 + a_2\Sigma X^3 \\ \Sigma X^2Y = a_0\Sigma X^2 + a_1\Sigma X^3 + a_2\Sigma X^4 \end{cases}$$

SERIES CRONOLÓGICAS

Las series de tiempo llamadas también series cronológicas o series históricas son un conjunto de datos numéricos que se obtienen en períodos regulares y específicos a través del tiempo, los tiempos pueden ser en años, meses, semanas, días u otra unidad adecuada al problema que se esté trabajando.

Ejemplos de series de tiempo son: Ventas mensuales de un producto en una empresa, producción total anual de petróleo en Ecuador durante un cierto número años o las temperaturas anunciadas cada hora por el meteorólogo para un aeropuerto.

Matemáticamente, una serie de tiempo se define por los valores $Y_1, Y_2, Y_{3,}$ .... de una variable Y (ventas mensuales, producción total, etc.) en tiempos $t_1, t_2, t_{3,}$ .... Si se reemplaza a X por la variable tiempo, estas series se definen como distribuciones de pares ordenados (X,Y) en el plano cartesiano, siendo Y una función de X; esto se denota por:

$$Y = f(t) \rightarrow Y = f(X)$$

El principal objetivo de las series de tiempo es hacer proyecciones o pronósticos sobre una actividad futura, suponiendo estables las condiciones y variaciones registradas hasta la fecha, lo cual permite planear y tomar decisiones a corto o largo plazo. Después, con base en esa situación ideal, que supone que los factores que influyeron en la serie en el pasado lo continuarán haciendo en el futuro, se analizan las tendencias pasadas y el comportamiento de las actividades bajo la influencia de ellas; por ejemplo, en la proyección de ventas de un producto o de un servicio de una empresa se calculan los posibles precios, la reacción del consumidor, la influencia de la competencia, etc.

MOVIMIENTOS O COMPONENTES DE SERIES DE TIEMPO

El modelo clásico o de descomposición, considera que los datos de series de tiempo están compuestas de los siguientes cuatro patrones básicos:

Tendencia secular.-

La tendencia secular o simplemente tendencia, son movimientos o variaciones continuas de la variable de modo uniforme y suave, por encima o por debajo, que se observan en el largo plazo durante un período de longitud prolongada.

Representan el comportamiento predominante o dirección general de la serie de tiempo como ascendente o descendente. La gráfica de la tendencia suele ser una curva suave y aun una línea recta que muestra la tendencia de las variaciones. Ejemplos de tendencia secular son las ventas, exportaciones, producción y el empleo.

La siguiente gráfica muestra la tendencia de exportaciones de la Empresa D & M en período 2000-2009. Aunque los datos muestran ciertas variaciones están por encima y por debajo de la recta de tendencia, la tendencia secular es ascendente.

Movimientos estacionales.-

Representa un movimiento periódico que se producen en forma similar cada año por la misma época, en correlación con los meses o con las estaciones del año y aun con determinadas fechas. Si los sucesos no se repiten anualmente, los datos deben recolectarse trimestral, mensual o incluso semanalmente.

Ejemplos de movimientos estacionales son la variación de precios de ciertos productos, incremento de ventas de juguetes y disminución de ventas de útiles Navidad, incremento de ventas de flores por el día del amor y la amistad, etc.

Movimientos cíclicos.-

Son variaciones hacia arriba y hacia abajo de la tendencia que se presentan cada cierto número de intervalos, en forma periódica de manera ondular a modo de oscilaciones más o menos regulares durante un período relativamente prolongado, que por lo general abarca tres o más años de duración. La producción, empleo, promedio industrial, etc. son ejemplos de este tipo de movimientos.

Movimientos irregulares o aleatorios.-

Son aquellas variaciones producidas por sucesos de ocurrencia imprevisible o accidental que producen movimientos sin un patrón discernible; así por ejemplo, las exportaciones de una empresa pueden ser afectadas por sucesos inusuales no previsibles tales como huelgas, guerras, terremotos, inundaciones, etc. Estas variaciones irregulares son de corta duración y de magnitud muy variable.

MODELOS DE SERIES DE TIEMPO

Son expresiones matemáticas de relación entre los movimientos de tendencia secular (T), movimientos cíclicos (C), movimientos estacionales (E) y movimientos irregulares (I) que generan la variable Y. Hay dos modelos para la definición de Y, los cuales son:

Modelo multiplicativo.-

En el que Y queda definida por el producto de las variaciones.

$$Y = T \cdot C \cdot E \cdot I$$

Modelo aditivo.-

En el que Y queda definida por la suma de las variaciones.

$$Y = T + C + E + I$$

En el modelo multiplicativo, las variaciones se expresan en términos relativos o porcentuales de la tendencia, en tanto que en el modelo aditivo las variaciones se expresan como residuos en las mismas unidades originales. El modelo aditivo sufre el supuesto irreal de que los movimientos o componentes son independientes uno de otro, algo que difícilmente se da en el caso de la vida real. El modelo multiplicativo supone que los movimientos o componentes interactúan entre sí y no se mueven independientemente, por lo que este modelo es más utilizado que el aditivo. Sin embargo, el criterio fundamental que se debe seguir en el caso de una situación dada es emplear el modelo que mejor se ajuste a los datos.

MÉTODOS DE SUAVIZAMIENTO Y PRONÓSTICO.-

Estos métodos eliminan las fluctuaciones aleatorias de la serie de tiempo, proporcionando datos menos distorsionados del comportamiento real de misma *(método de los promedios móviles y suavización exponencial)*

ANÁLISIS DE TENDENCIA.-

Es necesario describir la tendencia ascendente o descendente a largo plazo de una serie cronológica por medio de alguna línea, y la más adecuada será la que mejor represente los datos y sea útil para desarrollar pronósticos. Para lograr la estimación de la tendencia se utilizan con más frecuencia los siguientes métodos: *Método de los mínimos cuadrados y método de los semipromedios*

AUTOEVALUACIÓN

Con sus propias palabras y de manera lógica, defina cada uno de los siguientes conceptos:

1-. Defina correlación y regresión

2-. Análisis de correlación

3-. Coeficientes de correlación

4-. Coeficiente de determinación

5-. Análisis de regresión

6-. La recta de los mínimos cuadrados

7-. La parábola de los mínimos cuadrados

8-. Series cronológicas

9-. Movimientos o componentes de series de tiempo

10-. Tendencia secular

11-. Movimientos estacionales

12-. Movimientos cíclicos

13-. Movimientos irregulares o aleatorios

14-. Modelos de series de tiempo

15-. Modelo multiplicativo

16-. Modelo aditivo

17-. Métodos de suavizamiento y pronóstico

18-. Análisis de tendencia.-

GRÁFICOS Y TABLAS

GRÁFICOS ESTADÍSTICOS BÁSICOS

Las empresas, industrias, instituciones, etc. emplean diversos gráficos estadísticas para presentar informaciones sobre diversos asuntos relativos a ellas.

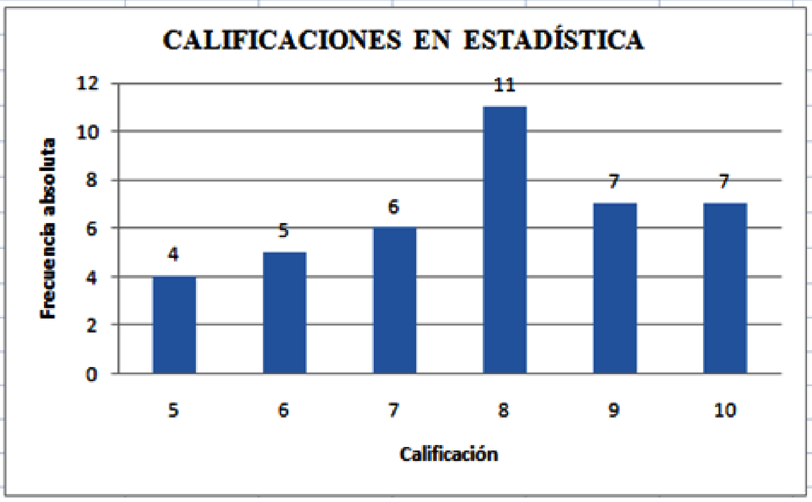
Las representaciones gráficas deben conseguir que un simple análisis visual ofrezca la mayor información posible. Según el tipo del carácter que estemos estudiando, usaremos una representación gráfica u otra.

A continuación se presenta los diagramas más empleados:

DIAGRAMAS DE BARRAS.-

Es un gráfico bidimensional en el que los objetos gráficos elementales son rectángulos de igual base cuya altura sea proporcional a sus frecuencias. Si en el eje horizontal se ubican las etiquetas con los nombres de las categorías, y en el eje vertical la frecuencia absoluta, la relativa o la frecuencia porcentual, toma el nombre de diagrama de barras vertical, y si se intercambian las ubicaciones de las categorías y las frecuencias, toma el nombre de diagrama de barras horizontal.

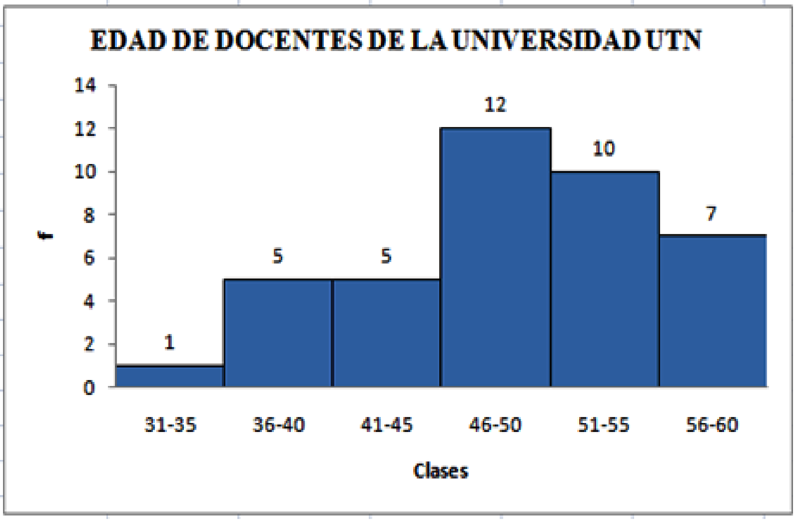
Calificación	f
5	4
6	5
7	6
8	11
9	7
10	7
Total	40



HISTOGRAMAS.-

Se utiliza para datos agrupados en intervalos de clase, representando en el eje horizontal los intervalos de clase o la marca de clase, y en el eje vertical se elabora rectángulos contiguos de base el ancho del intervalo y de altura proporcional a las frecuencias representadas.

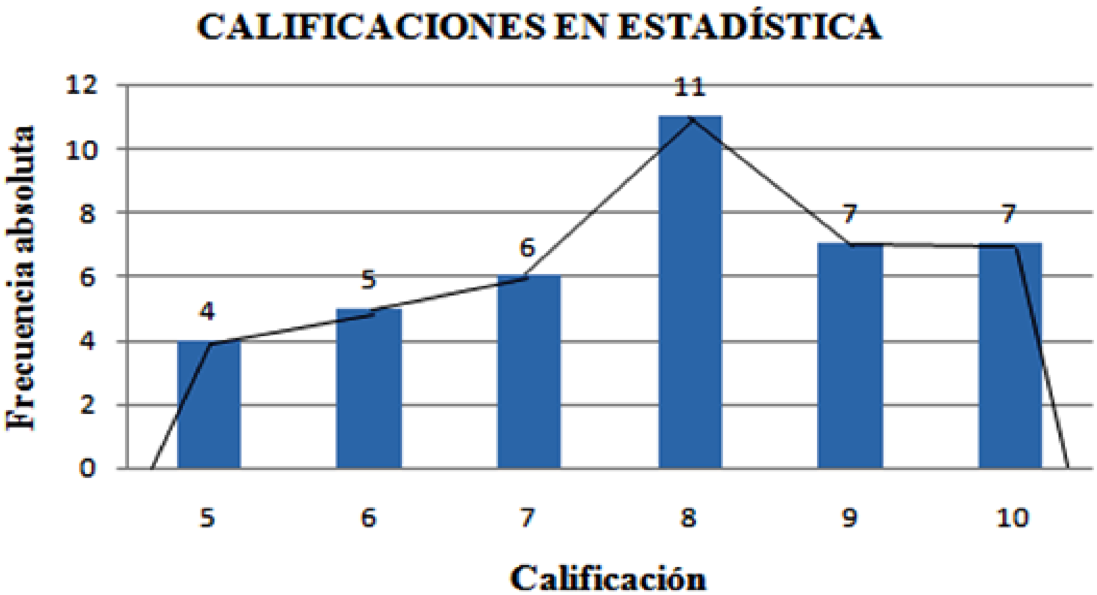
Clases	f
31-35	1
36-40	5
41-45	5
46-50	12
51-55	10
56-60	7
Total	40



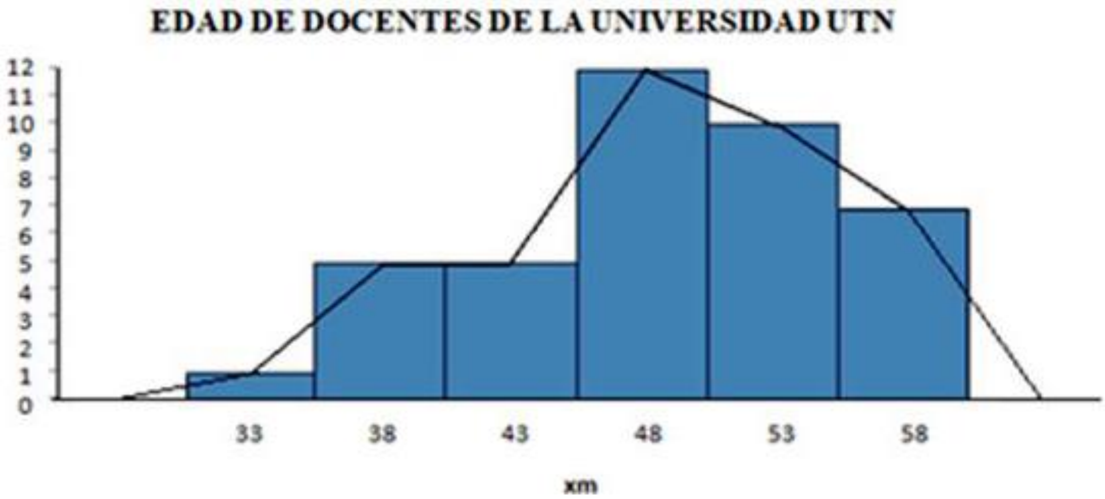
POLÍGONO DE FRECUENCIAS

Son gráficos lineales que se realizan uniendo:

a) Los puntos medios de las bases superiores de los rectángulos en un diagrama de barras.



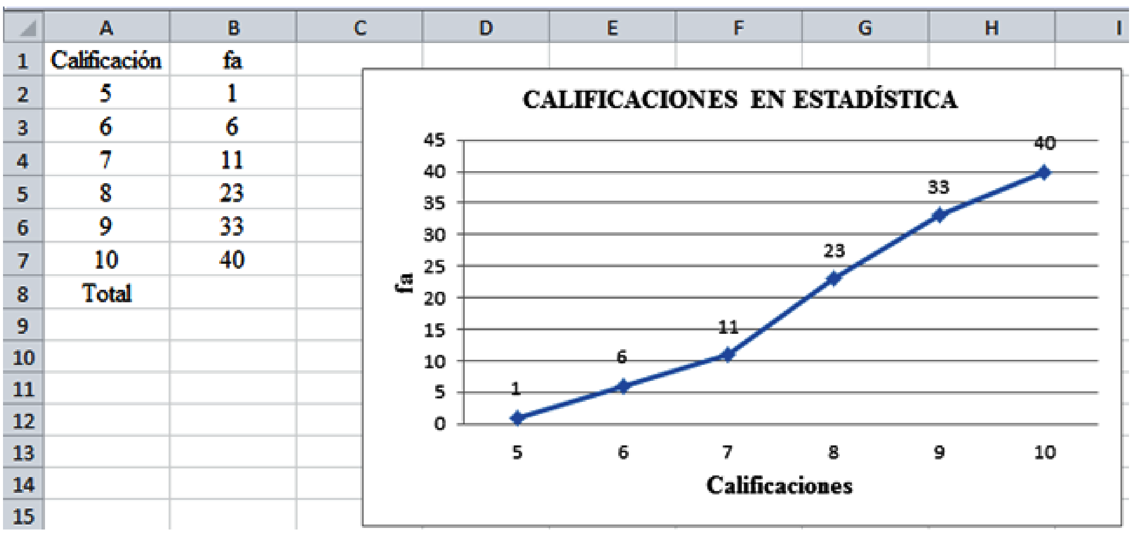
b) Los puntos medios (marcas de clase) de las bases superiores en el histograma.



Polígono de Frecuencias Acumuladas u Ojiva.-

Un gráfico que recoja las frecuencias acumuladas por debajo de cualquiera de las fronteras de clase superiores respecto de dicha frontera se llama un polígono de frecuencias acumuladas u ojiva.

Empleando polígono de frecuencias en 2D anterior, borrando la columna de la frecuencia absoluta y escribiendo la columna de la frecuencia acumulada del ejemplo del cálculo de las frecuencias sobre las siguientes calificaciones obtenidas por 40 estudiantes en una evaluación de la asignatura de Estadística se obtiene la siguiente figura que representa a una Ojiva:



Polígono de Frecuencias Relativas Acumuladas Porcentuales.-

Si se usan frecuencias fra% para realizar un polígono de frecuencias, este recibe el nombre de polígono de frecuencias relativas acumuladas porcentuales, o también llamado o*jiva de porcentajes*.

A continuación se presenta una ojiva de porcentajes elaborada en Excel empleando los datos del ejemplo de la Edad de 40 Docentes de la Universidad UTN:

Clases	xm	fra%
31-35	33	2,5
36-40	38	15
41-45	43	27,5
46-50	48	57,5
51-55	53	82,5
56-60	58	100
Total		

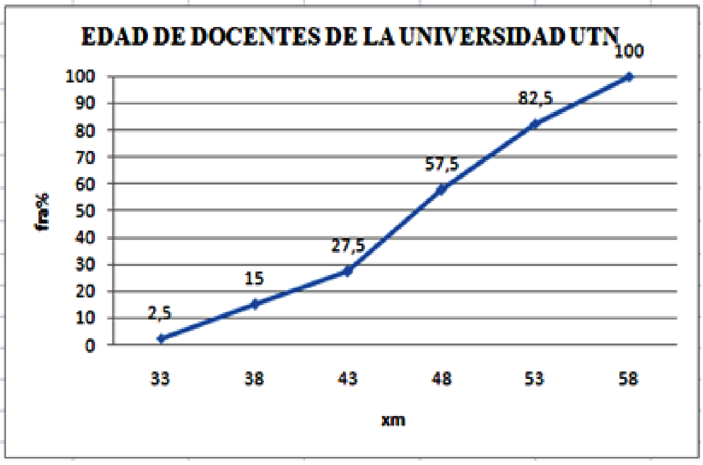


DIAGRAMA DE TALLO Y HOJAS

En el diagrama de tallo y hojas cada dato representa su valor y, a la vez, ocupa un espacio de forma que se obtiene simultáneamente la presentación de los datos y distribución gráfica.

En este diagrama cada valor se descompone en 2 partes: el primero o primeros dígitos (el tallo) y el dígito que sigue a los utilizados en el tallo (las hojas).Por ejemplo, el valor 32 puede descomponerse en un tallo de 3 y una hoja de 2; el valor 325 puede descomponerse en un tallo de 32 y una hoja de 5; el valor 3256 puede descomponerse en un tallo de 325 y una hoja de 6. Cada tallo puede ocupar una o más filas. Si un tallo ocupa una sola fila, sus hojas contendrán dígitos del 0 al 9; si ocupa dos filas, la primera fila contendrá dígitos del 0 al 4 y la segunda fila del 5 al 9.

La ventaja de este diagrama es que refleja a primera vista las mismas impresiones gráficas que el histograma sin necesidad de elaborar el gráfico. También tiene la ventaja de conservar los valores originales de los datos.

Ejemplo ilustrativo:

A 40 estudiantes se les pidió que estimen el número de horas que habrían dedicado a estudiar la semana pasada (tanto en clase como fuera de ella), obteniéndose los siguientes resultados:

30	30	32	32	35	35	35	35
36	37	38	39	39	40	45	45
47	47	47	48	48	49	50	50
50	52	54	55	55	56	56	56
58	58	58	58	58	60	60	65

Solución:

A fin de elaborar el diagrama de tallo y hojas se ordena los datos con los dígitos iniciales de cada uno, las decenas (tallos) a la izquierda de una línea vertical, y a la derecha de esa recta el último dígito de cada dato, en este caso la unidad, conforme recorren los datos en el orden en que fueron anotados.

3	0022
3	555567899
4	0
4	55777889
5	00024
5	5566688888
6	00
6	5

Interpretaciones:

Hay 4 estudiantes que dedican entre 30 y 32 horas semanales a estudiar, 10 estudiantes que dedican entre 55 y 58 horas semanales a estudiar, existe un solo estudiante que dedica 65 horas semanales a estudiar.

DIAGRAMA DE SECTORES

Llamado también diagrama circular o de pastel. Es un gráfico en el que a cada valor o modalidad se asigna un sector circular de área proporcional a la frecuencia que representan.

Ejemplo ilustrativo:

Con los datos de la siguiente tabla sobre las calificaciones obtenidas por 40 estudiantes en una evaluación de Estadística, presentar la información a través de un diagrama de sectores:

X_i	f_i
5	4
6	5
7	6
8	11
9	7
10	7
Total	40

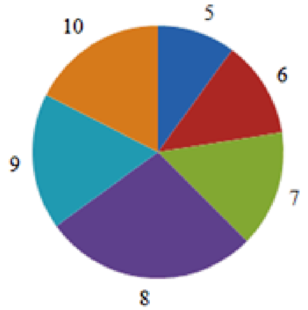
Solución:

a) Se calcula la frecuencia relativa y el número de grados que representa cada calificación. El número de grados se calcula multiplicando la frecuencia relativa con 360°, así: número de grados = fr * 360°

Estos cálculos se muestran en la siguiente tabla:

Xi	**f**	**fr**	**°**
5	4	0,100	36
6	5	0,125	45
7	6	0,150	54
8	11	0,275	99
9	7	0,175	63
10	7	0,175	63
Total	**40**	**1**	**360**

b) Se dibuja una circunferencia tomando para cada calificación tantos grados como indica la tabla anterior como se muestra en la siguiente figura:



PICTOGRAMAS

Son dibujos, figuras o signos llamativos alusivos al carácter que se está estudiando cuyo tamaño es proporcional a la frecuencia que representa los datos.

Ejemplo ilustrativo:

Un equipo de fútbol en su trayectoria tiene 120 partidos ganados, 60 perdidos y 30 empatados. Al representar estos datos mediante pictogramas se obtiene:

Otra forma de representar los datos mediante pictogramas se muestra en la siguiente figura:

AUTOEVALUACIÓN

Defina cada uno de los siguientes gráficos y elabore ejemplos de cada uno:

1-. Diagramas de barras

2-. Gráficos estadísticos básicos

3-. Histogramas

4-. Polígono de frecuencias

5-. Polígono de frecuencias acumuladas u ojiva

6-. Polígono de frecuencias relativas acumuladas porcentuales

7-. Diagrama de tallo y hojas

8-. Diagrama de sectores

9-. Pictogramas

REFERENCIAS BIBLIOGRÁFICAS

Levin, Richard y Rubin, David *(2004) Estadística para Administración y Economía.* Editorial Pearson Educación.

Silva Jesús Alirio (2006) *Metodología de la investigación. Elementos básicos*. Ed. Colegial Bolivariana. Caracas. Venezuela.

Sweeny, Denis y Thomas Willians (2008) *Estadística para administración y economía* Edit. Internacional Thomson

Pestaña de Martínez, Pilar: (2002) *Estadística: conceptos básicos, terminología y metodología de la estadística descriptiva. Los libros de El Nacional, CEC, S.A*

Webster A. M. (2002) *Estadística aplicada a los negocios y a la economía* Edit. Mc Graw Hill.