Estadística

❖ ¿Por qué hay que estudiar Estadística?

Si se revisa un catálogo de información de universidad, se descubrirá que la educación estadística se requiere en muchos programas escolares. ¿Porqué pasa esto?. ¿Cuáles son las diferencias en los cursos de Estadística impartidos en una Facultad de Ingeniería, en Departamentos de Psicología o Sociología de una universidad, y los de un instituto o Escuela de Administración?. La mayor diferencia son los ejemplos utilizados.

Básicamente, el contenido del curso es el mismo; en una Escuela de Administración interesan cosas como las ganancias, horas de trabajo, y salarios. En un Departamento de Psicología interesan los resultados de las pruebas, y en una Facultad de Ingeniería puede interesar cuántas unidades son producidas por una máquina en especial. Sin embargo, las tres áreas tienen interés en lo que es un valor típico y en la cantidad de variación existente en la información. Es posible que también exista una diferencia en el nivel de matemáticas requerido. Un curso de Estadística en ingeniería generalmente requiere del Cálculo, los cursos de Estadística en escuelas de administración y en la educación, generalmente enseñan un curso orientado a aplicaciones. Entonces, ¿por qué se requiere estudiar Estadística en tantas carreras?.

La primera razón es que en todos lados encontramos información numérica. Si se revisan los periódicos, revistas de información, revistas de negocios, publicaciones de interés general, o revistas de deportes, uno estará bombardeado con información numérica.

Presentamos aquí algunos ejemplos:

- Ford reporta que en 1996 sus ventas fueron de $146900 millones (de dólares), arriba en un 7,2%; sus ganancias fueron de $4400 millones, con ascenso en un 7,0%, y el efectivo neto circulante fue de $7200 millones.
- Los egresados de postgrado del Programa de Maestría en Administración de Empresas en la Universidad de Notre Dame, contaron con un sueldo promedio inicial de $54000 dólares y un 91% de ellos consiguieron trabajo a los tres meses de la graduación.
- Para los golfistas que gustan de jugar en campos de golf públicos, las cuotas de los campos promediaban $176,20 dólares por año.

¿Cómo podemos determinar si las conclusiones presentadas son razonables?, ¿las muestras fueron suficientemente grandes?, ¿cómo se seleccionaron las unidades de la muestra?. Para poder ser un consumidor con conocimientos sobre esta información, necesitamos poder leer los cuadros, las gráficas y entender la discusión de la información numérica. El entender los conceptos básicos de la Estadística será de gran ayuda.

La segunda razón para tomar el curso de Estadística es que las técnicas estadísticas se utilizan para tomar decisiones que afectan nuestra vida diaria. Esto quiere decir que afectan a nuestro bienestar personal. He aquí algunos ejemplos:

- Las compañías de seguros utilizan análisis estadísticos para establecer las tarifas de los seguros de casa, automóvil, vida y salud. Existen tablas que resumen la probabilidad de que una mujer de 25 años de edad viva el año siguiente, los siguientes cinco años, etc. Las primas del seguro de vida se pueden establecer basándose en estas probabilidades.
- La Agencia de Protección al Medio Ambiente está interesada en la calidad del agua en el Lago Ene. Periódicamente toman muestras de agua para establecer el nivel de contaminación y mantener el nivel de calidad.

- Los investigadores médicos estudian las tasas de cura de enfermedades, basándose en el uso de diferentes medicamentos y distintas formas de tratamiento. Por ejemplo, ¿cuál es el efecto de tratar cierto tipo de daño a la rodilla con cirugía o con terapia física?. Si se toma una aspirina diaria, ¿se reducirá el riesgo de sufrir un ataque cardiaco?.

La tercera razón para tomar el curso de Estadística es que el conocimiento de los métodos estadísticos ayudarán a entender por qué se toman ciertas decisiones, y le aportarán una mejor comprensión sobre la manera en la que lo afectan.
Sin importar el tipo de trabajo que seleccione, encontrará que tiene que enfrentar la toma de decisiones con la ayuda del análisis de datos. Para poder realizar una decisión basada en la información, necesitará:

1. Determinar si la información existente es adecuada o si se requiere información adicional.
2. Reunir información adicional, si es necesario, de tal forma que no hayan resultados erróneos.
3. Resumir la información de una forma útil e informativa.
4. Analizar la información disponible.
5. Sacar las conclusiones y realizar las deducciones necesarias, al tiempo que se evalúa el riesgo de llegar a una conclusión incorrecta.

❖ Definición

La estadística es la ciencia destinada al estudio de los fenómenos aleatorios, la misma está ligada con los métodos científicos en la toma, recopilación, organización, presentación y análisis de datos; tanto para la deducción de conclusiones como para la toma de decisiones razonables de acuerdo con tales análisis.

❖ Clasificación

1) Estadística Descriptiva: cuando se describe, analiza y representa un grupo de datos utilizando métodos numéricos y gráficos que resumen y presentan la información contenida en ellos. No pretende ir más allá del conjunto de datos investigados.

2) Estadística Inferencial: cuando apoyándose en el cálculo de probabilidades y a partir de datos muestrales, efectúa estimaciones, decisiones, predicciones y otras generalizaciones sobre un conjunto mayor de datos.

❖ Elementos que caracterizan a los problemas estadísticos

1) La población de interés y el procedimiento científico que se empleó para tomar la muestra de la población.
2) La muestra y el análisis matemático de su información.
3) Las inferencias estadísticas que resultan del análisis de la muestra.
4) La probabilidad de que las inferencias sean correctas.

❖ Definiciones básicas

1) Datos: son los hechos, medidas o números que han sido recopilados como resultados de observaciones; se deben reunir, analizar y resumir para su presentación e interpretación. Pueden ser cuantitativos (siempre numéricos) o cualitativos (que pueden ser numéricos o no, ya que son etiquetas o nombres asignados a un atributo de cada elemento. Por ejemplo: el sexo de una

persona es masculino o femenino, pero podrían ser codificadas con 1 o 2 y en este caso, los números sólo servirían para indicar la categoría y no tendrían significación numérica).

2) Toma de datos: es la obtención de una colección de datos que no han sido ordenados numéricamente. Ejemplo: el conjunto de estaturas de 100 estudiantes, sacados de una lista alfabética de una universidad.

3) Individuos o elementos: seres u objetos que contienen cierta información que se desea estudiar.

4) Población (N): es el conjunto de todas las observaciones o de los elementos de interés en un determinado estudio que cumplen ciertas propiedades comunes. Este conjunto puede ser un número finito de datos o una colección grande (virtualmente infinita) de datos.
Por ejemplo, se puede considerar como una población finita a todas las ruedas fabricadas por la Goodyear en un año, mientras que el conjunto de todos los resultados posibles al lanzar una moneda de forma sucesiva constituye una población infinita.

5) Parámetro: es cualquier medida descriptiva de una población, por ejemplo, la media poblacional.

6) Muestra (n): es un subconjunto de la población, sin embargo, nos interesa que ese subconjunto seleccionado de la población sea representativo, esto significa que debe contener las características relevantes de la población en la misma proporción en que están incluidas en dicha población. Las muestras pueden ser probabilísticas (aleatoria simple, estratificadas, por conglomerados, etc.) o no probabilísticas (por juicio, por cuota, etc.).

7) Estadístico: es cualquier medida descriptiva de una muestra y se usa como base para estimar el parámetro correspondiente de la población. Por ejemplo, la media muestral.

8) Caracteres: son propiedades, rasgos o cualidades de los elementos de la población. Estos caracteres pueden dividirse en cualitativos y cuantitativos. Tradicionalmente a los caracteres cualitativos se les ha llamado *atributos* y a sus distintas formas de presentación modalidades, mientras que los cuantitativos han recibido el nombre de *variables* y los posibles resultados de sus observaciones valores. A menos que se especifique lo contrario, se utilizará la expresión de *variable* como nombre genérico para la descripción de cualquier tipo de carácter .

9) Variable: es un carácter de la muestra o de la población que se observa. Entre los tipos de variable tenemos:
V. cualitativa: cuando la característica de estudio es no numérica; por ejemplo: la preferencia religiosa, el sexo, el color del cabello, el estado civil, etc.
V. cuantitativa: es aquella que asume valores numéricos acompañados de una unidad de medida; por ejemplo: calificaciones de un examen.
V. continua: es aquella que puede tomar cualquier valor dentro de un intervalo, por lo general los valores de una variable continua proceden de mediciones. Ejemplos: la estatura, el tiempo en realizar una transacción bancaria, la presión de aire en un caucho, etc.
V. discreta: es aquella que sólo puede tomar determinados valores en un intervalo, por lo general son números enteros, y suelen ser el resultado de un conteo. Ejemplo: el número de hijos de una familia, el número de habitaciones de una casa, etc.

10) Fuentes para la recolección de datos: a fin de que un análisis estadístico resulte útil en el proceso de toma de decisiones, los datos de entrada iniciales deben ser apropiados ya que si son ambiguos o tienen algún tipo de error, es posible que no se puedan compensar estas deficiencias. Son variados los métodos que pueden utilizar los investigadores para obtener los datos necesarios para su estudio, entre estos tenemos:

- Buscar datos publicados por fuentes gubernamentales, industriales o particulares.
- A través del diseño de un experimento.
- A través de encuestas, entrevistas, cuestionarios, etc.
- Internet.

❖ Razones, proporciones y porcentajes

Una de las funciones de los métodos estadísticos es la de resumir todos los datos de una serie de valores, para poner de manifiesto las características más importantes de dicha serie. La forma más simple de cumplir esta función es convertir los datos de valores absolutos en relativos, esta conversión se hace necesaria debido a que los valores relativos pueden contener todas las informaciones que interesan, lo que no se logra con los absolutos (como para la comparación de dos poblaciones de cantidades de diferentes unidades). Para ello debemos conocer el significado de razón, proporción y porcentaje.

1) Razón: es aquel valor que indica la relación cuantitativa existente entre dos cantidades.

$$R = \frac{\text{número de individuos que poseen cierta característica}}{\text{número de individuos que no poseen dicha característica}}$$

Ejemplo:

Si en una determinada zona existen 32000 empleados y 8000 desempleados, la razón de empleado a desempleado viene dada por:

La característica viene dada por el hecho de estar empleado, luego:

$$\frac{32000}{8000} = \frac{4}{1} \Rightarrow \text{por cada 4 empleados hay 1 desempleado.}$$

2) Proporción: es una razón, en la cual el denominador es el número total de unidades enunciadas. Siguiendo con el ejemplo anterior:

$$\text{Proporción de empleados:} \quad \frac{32000}{40000} = 0{,}80$$

$$\text{Proporción de desempleados:} \frac{8000}{40000} = 0{,}20$$

3) Porcentaje: se llama tanto por ciento de un número a una o varias de las 100 partes iguales en que se puede dividir dicho número. Por ejemplo, el 4% de 80, significa que el 80 se divide en 100 partes iguales y de ellas se toman 4. También es una medida que se obtiene al multiplicar por 100 a las proporciones.

Casos:

1) Hallar un tanto por ciento de un número:

¿Cuál es el 15% de 32?

$$\begin{array}{l} 32 - 100\% \\ \;x - 15\% \end{array} \qquad \text{Luego } x = 4{,}8$$

2) Hallar un número cuando se conoce un tanto por ciento de él:

¿De qué número es 46 el 23%?

$$\begin{array}{l} 46 - 23\% \\ x - 100\% \end{array} \quad \text{Luego } x = 200$$

3) Dados 2 números, determinar qué tanto por ciento es uno del otro:

¿Qué porcentaje de 8400 es 2940?

$$\begin{array}{l} 8400 - 100\% \\ 2940 - x \end{array} \quad \text{Luego } x = 35\%$$

4) Tanto por ciento más y tanto por ciento menos:

¿De qué número es 265 el 6% más?

$$\begin{array}{l} 265 - 106\% \\ x - 100\% \end{array} \quad \text{Luego } x = 250$$

¿De qué número es 168 el 4% menos?

$$\begin{array}{l} 168 - 96\% \\ x - 100\% \end{array} \quad \text{Luego } x = 175$$

Distribuciones de Frecuencia

Una vez que se han recogido y tabulado los datos, los mismos deben ser presentados de una manera organizada para facilitar el acceso a la información que contienen. Ahora bien, si el conjunto de datos es grande, la mejor manera de examinar estos datos es presentarlos en forma resumida, elaborando las tablas y gráficas apropiadas, de esta forma se pueden extraer las principales características de los datos.
Aunque en el proceso de agrupamiento generalmente se pierde parte del detalle original de los datos, tiene la importante ventaja de presentarlos a todos en un sencillo cuadro que facilita asimilar la información. Como podemos manejar diversos tipos de datos, empezaremos primero con datos cuantitativos y luego con los cualitativos.

- **Para datos cuantitativos**

Para el manejo de datos cuantitativos, se establecerán los siguientes conceptos:
1) Frecuencia Absoluta (f_i): es un número que indica la cantidad de veces que se repite un dato.

2) Distribución de frecuencias: es un resumen tabular de un conjunto de datos que muestra la cantidad de elementos en cada una de las diferentes clases que la conforman.

3) Distribución de frecuencias para datos no agrupados: es una tabla compuesta por dos columnas, en una se ubican los valores de la variable y en otra sus respectivas frecuencias absolutas. Ejemplo:
Supongamos que los siguientes datos corresponden al peso en Kg. de un grupo de estudiantes: 56, 58, 61, 62, 67, 68, 70, 75, 56, 58, 61, 68, 75, 58, 68, 68. Al construir la tabla de distribución de frecuencias obtenemos:

Peso (Kg.)	f_i
56	2
58	3
61	2
62	1
67	1
68	4
70	1
75	2

4) Distribución de frecuencias para datos agrupados: es una tabla resumen en la cual los datos se encuentran divididos en grupos ordenados numéricamente. A estos grupos se les denominan clases o categorías.

5) Pasos para la construcción de la tabla de distribución de frecuencias para datos agrupados:

1. Selección del número de clases: el número de clases que se utilizan depende primordialmente de la cantidad de datos que se tengan, es una decisión arbitraria; sin embargo, en términos generales, se recomienda que la distribución de frecuencias deba tener al menos 5 clases y no más de 15 (si no existen suficientes clases, o si hay demasiadas, la información que se puede obtener es reducida). Entre las expresiones que se pueden utilizar para calcular el número de clases tenemos:

- $d \geq \frac{\text{Log n}}{\text{Log 2}}$ donde "d" es el número de clases y "n" el número total de observaciones.
- $\sqrt{n}$ donde "n" el número total de observaciones.
- 1+3,322 Log n (regla de Sturges) y "n" el número total de observaciones.

Estas reglas no deben tomarse como un factor determinante o definitivo, por ejemplo, si el número de observaciones que tenemos es 100, es un buen criterio agrupar las observaciones en $\sqrt{100}$ = 10 intervalos, pero si el número de observaciones fuese muy alto como por ejemplo n = 1000000, este segundo criterio nos da un número excesivo de intervalo (1000) por lo que en estos casos habrá que hacer uso del sentido común para determinar el número de intervalos.

2. Obtención de los intervalos de clase: el intervalo de clase es el recorrido de los valores que se encuentran dentro de una clase, es recomendable al elaborar la tabla que todas las clases tengan el mismo tamaño porque facilita la interpretación estadística de cualquier utilización posterior que se pueda hacer de los datos.
   Mediante la expresión:
   $$\text{Tamaño del intervalo} = \frac{\text{medida mayor - medida menor}}{\text{número de clases que se desean}}$$
   Obtenemos un valor que sirve de guía para establecer el tamaño de los intervalos, el valor numérico que obtengamos de la fórmula anterior lo podemos redondear dependiendo de nuestra conveniencia, pero en cualquier caso, se toma con un grado de aproximación no mayor a aquel con el que se registran los datos.
   Si todos los intervalos de clase tienen el mismo tamaño, representaremos con la letra "a" al tamaño del intervalo y lo llamaremos amplitud del intervalo.

   A manera de información: aunque anteriormente se recomendó que todas las clases tengan el mismo tamaño, existen casos donde esta regla no puede o no debe aplicarse; por ejemplo, si se tuviera a mano la lista de impuestos pagados por la población en un año, estas cantidades (supuestas) pueden encontrarse en un intervalo de Bs. 0 a Bs. 10000000, aún a pesar de que se eligiesen 20 clases para la distribución de frecuencia, con intervalos de igual longitud, cada clase tendría una cobertura de Bs. 500000. Lo anterior daría origen a una situación en la que casi todas las observaciones caerían en la primera clase, en casos como este, es preferible seleccionar una escala más pequeña en el extremo inicial que la utilizada para el extremo superior. También sería posible reducir el número de clases que se requieren cuando unos cuantos de los valores son mucho menores o mucho mayores que el resto, mediante ***clases abiertas*** (aunque se deben evitar cuando sea posible ya que dificultan calcular ciertas medidas o ciertas descripciones adicionales que puedan ser de interés.)
   Ejemplos de clases abiertas:

10 o Menos	10 o Menos	11 a 15
11 a 15	11 a 15	16 a 20
16 a 20	16 a 20	21 a 25
21 o Más	21 a 25	26 o Más

En la primera columna, es abierta en ambos extremos; en la segunda columna, es abierta en el extremo inferior y en la última, es abierta en el superior.

3. Establecimiento de los límites de clase: para construir la tabla de distribución de frecuencias se necesitan establecer límites claramente definidos para cada una de las clases, de manera que se eviten problemas como:
   - El solapamiento entre clases (no debe existir duda en la ubicación de los datos en las clases).
   - Que no se incluyan a todas las observaciones.

   Por ejemplo:
   Supongamos que se tienen un conjunto de datos (entre ellos está el valor de 60 y el de 70) los cuales se deciden agruparlos, dando como resultado los siguientes intervalos:

   $50-60$
   $60-70$ Se presenta el solapamiento. ¿A qué clase pertenece el 60 o el 70?
   $70-80$
   $\vdots$

   $51-59$
   $61-69$ Se presenta la exclusión de valores. ¿Ninguna clase contiene al 60 o al 70?
   $71-79$
   $\vdots$

   Los valores correspondientes a la primera columna son los límites inferiores. El límite inferior (l_i) se define como el valor mínimo posible de los datos que se asignan a la clase.
   Los valores correspondientes a la segunda columna son los límites superiores (l_{i+1}).
   Observaciones:
   - Una forma de obtener la amplitud de un intervalo es mediante la diferencia entre dos límites superiores consecutivos o dos límites inferiores consecutivos.
   - Es de hacer notar que la selección de los límites de clase es subjetiva y para conjuntos de datos que no contienen muchas observaciones, la selección de un conjunto específico de límites de clase y no otro, puede dar una imagen distinta al lector; sin embargo, al aumentar el número de observaciones de los datos, las alteraciones en la selección de los límites de clase afectan cada vez menos la concentración de los datos.
   - Algunos autores difieren en la forma en que toman los límites cuando construyen las tablas de distribución de frecuencias (básicamente según el tipo de variable con la cual trabaje), unos toman los intervalos de tal manera que son cerrados en el límite inferior y abiertos en el superior, entre los tipos de notación que se pudieran presentar tenemos:

$[50-60)$	50 a menos de 60	50 - <60
$[60-70)$	60 a menos de 70	60 - <70
$\vdots$	$\vdots$	$\vdots$

   Esta forma de notación es particularmente útil si se trabajan con variables continuas.

   Otra forma que toman los autores, es considerar los límites como *reales* o *fronteras de clase* e *imaginarios* o de *escritura*. Los límites reales son aquellos que reflejan la unidad más pequeña que se emplea para tomar las observaciones; los imaginarios son aquellos que reflejan el mismo grado de precisión que el de las observaciones presentadas. Para los efectos del curso, no se trabajará con este tipo de límites.

Ejemplo:
Los datos 23, 24, 18, 14, 20, 13, 38, 19, 16, 24, 11, 16, 18, 20, 23, 19, 32, 36, 15, 10, 20 son parte de un total de 80 datos que serán utilizados para construir una tabla de distribución de frecuencias, suponemos que los cálculos para determinar el número de clases y la amplitud ya fueron realizados dando como resultado que el número de clases es 6 y la amplitud es 5.

En la tabla se encuentran los datos que han sido agrupados:

$l_i - l_{i+1}$
[10 – 15)
[15 – 20)
[20 – 25)
[25 – 30)
[30 – 35)
[35 – 40)

Como observación: se debe estar pendiente que al tomar límites, los valores adecuados de los límites de clase con datos cuantitativos continuos dependen de la exactitud de los datos con los cuales se trabaja.

4. Establecimiento de la marca de clase (x_i): la marca de clase es un punto representativo del intervalo. Si éste es acotado tomamos como marca de clase al punto medio del intervalo (se asume que los valores de la variable se distribuyen de manera uniforme dentro del intervalo). Se obtiene como un promedio aritmético entre los límites superior e inferior de cada intervalo de clase.

$$x_i = \frac{l_{i+1} + l_i}{2}$$

6) Frecuencia relativa de clase (h_i): es el valor que se obtiene al dividir la frecuencia de clase entre el número total de observaciones, por lo que indica la proporción de la cantidad total de datos que pertenecen a una clase.

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

7) Distribución de frecuencias relativas: es una tabla donde se presentan las frecuencias relativas de clase.

8) Frecuencia acumulada de clase (F_i): es la frecuencia total de todos los valores que hasta su límite superior existen en la serie. Si trabajamos con intervalos cerrados y abiertos, para esta definición de frecuencia acumulada, no incluimos el valor del límite superior. Observación: esta definición es válida para distribuciones acumuladas “menor que”.

9) Distribución de frecuencias acumuladas "menor que": es una tabla donde se presentan las frecuencias acumuladas, para hallar esta distribución en una clase determinada lo que se hace es sumar la frecuencia de esa clase a la de las clases anteriores. Las distribuciones de frecuencias acumuladas nos permite ver cuántas observaciones se encuentran por arriba o debajo de ciertos valores.

10) Frecuencia relativa acumulada (H_i): es el cociente de la frecuencia acumulada con respecto a la frecuencia total, muestra la proporción de elementos con valores menores o iguales al límite superior de cada clase. Si trabajamos con intervalos cerrados y abiertos, para esta definición de frecuencia relativa acumulada, no incluimos el valor del límite superior

$$H_i = \frac{F_i}{n}$$

11) Distribución de frecuencias relativas acumuladas: es una tabla donde se presentan las frecuencias relativas acumuladas.

Cuando se pida construir una tabla de distribución de frecuencias consideraremos a todas las distribuciones anteriores.

Ejemplo:

Un investigador desea determinar cómo varían las estaturas de las obreras de una empresa y toma una muestra de 50 mujeres para registrar luego sus estaturas en pulgadas. Los datos obtenidos fueron los siguientes:

65	63	65	63	69	67	53	58	60	61
64	65	64	72	68	66	55	57	60	62
64	65	64	71	68	66	56	59	61	62
63	65	63	70	67	66	57	59	61	62
64	64	63	69	67	66	58	60	61	62

Construir la tabla de distribución de frecuencias:

1. Selección del número de clases:

   $\sqrt{n} = \sqrt{50} \approx 7{,}071$

   tomamos 7 clases.

2. Cálculo de la amplitud del intervalo:

   $a = \frac{72-53}{7} = 2{,}7$

   tomamos a = 3

3. Establecimiento de los límites y construcción de la tabla:

l_i - l_{i+1}	f_i	x_i	F_i	h_i	% h_i	H_i	% H_i
[53 – 56)	2	54,5	2	0,0400	4	0,0400	4
[56 – 59)	5	57,5	7	0,1000	10	0,1400	14
[59 – 62)	9	60,5	16	0,1800	18*	0,3200	32**
[62 – 65)	15	63,5	31	0,3000	30	0,6200	62
[65 – 68)	12	66,5	43	0,2400	24	0,8600	86
[68 – 71)	5	69,5	48	0,1000	10	0,9600	96
[71 – 74)	2	72,5	50	0,0400	4	1,0000	100
	n = Σ = 50			Σ =1	Σ =100		

Interpretaciones:

18* significa que el 18% de las obreras tienen estaturas que van desde 59 pulgadas, pero son menores de 62 pulgadas.

32** significa que el 32% de las obreras tienen una estatura inferior a 62 pulgadas.

❖ Para datos cualitativos

Hasta ahora sólo hemos analizado la construcción de distribuciones numéricas, pero el problema general que implica construir distribuciones cualitativas es casi el mismo. Una vez más debemos decidir cuántas clases utilizar y qué tipo de elementos contendrá cada categoría, asegurándonos que se puedan acomodar todos los datos y que no se presenten ambigüedades.

Como las categorías a menudo se escogen antes de que se recolecten los datos, es prudente incluir una categoría marcada con el título “otros” o “mixto”, la ventaja de trabajar con datos cualitativos es que no tenemos que preocuparnos por los límites de clase, las fronteras de clase, las marcas de clase, etc.

La construcción de una tabla de frecuencias para datos cualitativos requiere sólo del conteo del número de elementos o individuos que caen dentro de cierta clase o tienen determinada característica.

Ejemplo:
La siguiente tabla pertenece a los planes de estudios superiores de un grupo de 548 estudiantes del último año del bachillerato:

	f_i	h_i	$\%h_i$
Planean ir a la universidad	240	0,4379	43,79
Quizá vayan a la universidad	146	0,2703	27,03
Planea ir o quizá vayan a una escuela vocacional	57	0,1055	10,55
No irán a ninguna universidad	105	0,1944	19,44

Para la tabulación de datos cualitativos también se pueden usar tablas de contingencia o supertablas, el valor de una tabulación cruzada consiste en que proporciona una idea de la relación entre las variables (ya sean ambas cualitativas, ambas cuantitativas o combinación de ambas).

Ejemplo:
Un prestamista local tiene en la actualidad 120 cuentas, su contable le comunica que de las 25 cuentas comprendidas entre 0 y 4999 dólares; 10 vencen ahora, 5 vencieron hace tiempo y el resto son morosas; lo que implica para el deudor el peligro de ver ejecutada la deuda por el prestamista.

De las 37 cuentas situadas en el intervalo de 5000 a 9999 dólares; 15 vencen ahora, 10 han vencido hace tiempo y el resto son morosas.

Hay 39 cuentas en el intervalo de 10000 a 14999 dólares que indican que 11 vencen ahora, 10 vencieron hace tiempo y el resto son morosas. Del resto de las cuentas, en el intervalo de 15000 o más; 5 vencen ahora, 7 han vencido y el resto son morosas.

El prestamista quiere ver una tabla de contingencia de estas cuentas, para lo cual le pide a su contable que la elabore:

Cuentas / Condición	0 - 4999	5000 - 9999	10000 - 14999	15000 o más	Totales
Vencen ahora	10	15	11	5	41
Vencieron hace tiempo	5	10	10	7	32
Morosas	10	12	18	7	47
Totales	25	37	39	19	120

Gráficas de las distribuciones de frecuencia

La afirmación "una imagen vale más que mil palabras" se puede aplicar al ámbito de la estadística descriptiva diciendo que "un gráfico bien elaborado vale más que mil tablas de frecuencia". Cada vez es más habitual el uso de gráficos o imágenes para representar la información obtenida; de todas maneras, debemos ser prudentes al confeccionar o interpretar gráficos, puesto que una misma información se puede representar de formas muy diversas y no todas ellas son pertinentes, correctas o válidas.

- **Pictogramas**

Son presentaciones gráficas que se hacen por medio de dibujos, que en la mayoría de los casos son semejantes al fenómeno que se quiere representar. Por ejemplo, si se fuese a representar la población de un determinado estado clasificado por distritos, se identifica a esta población a través de figuras humanas; por medio de estos dibujos se expresan las frecuencias de las modalidades de la variable.
También estos gráficos se hacen representando en diferentes escalas un mismo dibujo, la escala de los dibujos debe ser tal que el área de cada uno de ellos sea proporcional a la frecuencia de la modalidad que representa.
Ejemplo:

Botellas de cerveza recogidas en un fin de semana en la calle

500 Botellas
Ciudad A

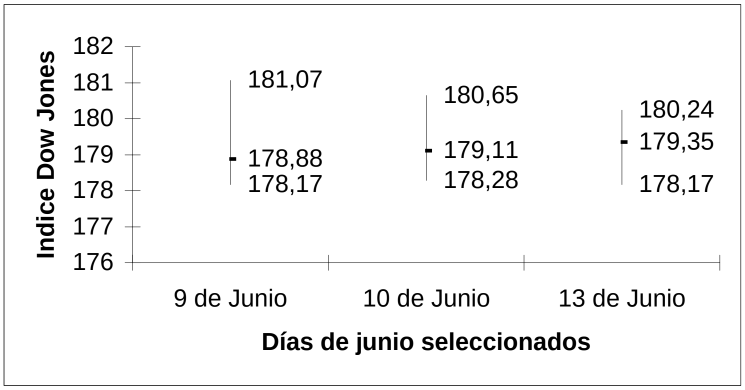
125 Botellas
Ciudad B

- **Gráfico de máximo-mínimo-al cierre**

Como su nombre lo indica, son gráficos que presentan el valor máximo, el mínimo y el último valor de una variable seleccionada durante un período determinado; el ejemplo quizá más conocido es el índice Dow Jones.
Ejemplo:

	Máximo	Mínimo	Al cierre
Junio 9	181,07	178,17	178,88
Junio 10	180,65	178,28	179,11
Junio 13	180,24	178,17	179,35

La gráfica será:



❖ Gráficos circulares, de sectores o de pastel

Este tipo de gráfico considera al círculo como la totalidad del fenómeno, en consecuencia, se dividirá al mismo en tantos sectores como componentes tenga el fenómeno a representar; son bastante útiles para visualizar diferencias de porcentajes, para representar datos cualitativos, etc.

Pasos para su construcción:

1. Buscamos los porcentajes que representan a cada elemento.
2. Cada porcentaje se multiplica por 3,6 y eso nos daría el valor de los ángulos centrales.
3. Utilizar un transportador para ubicar cada ángulo.

También es posible hallar los ángulos centrales estableciendo una regla de tres entre la totalidad del fenómeno (al cual le corresponden 360°) y la frecuencia de cada parte del fenómeno.

Ejemplo:

Para estudiar sus actitudes hacia aspectos sociales, a 1200 personas se les preguntó si se está gastando "muy poco", "más o menos de lo debido" o "demasiado" en programas de bienestar social. Trace un gráfico circular para desplegar los resultados que se muestran en la tabla siguiente:

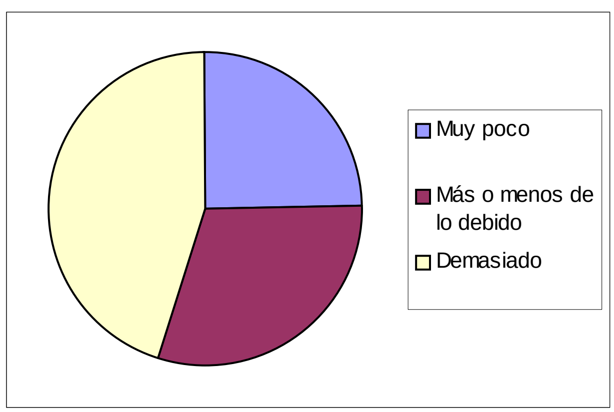
Opinión	f_i	h_i	% h_i
Muy poco	296	0,24666...	24,666
Más o menos de lo debido	360	0,3	30,0
Demasiado	544	0,45333...	45,333

Luego, los ángulos centrales serán:

24,666...*3,6 ≈ 88,8 grados

30,00*3,6 = 108 grados

45,333...*3,6 ≈ 163,2 grados



❖ Gráficos de tallo y hoja

Son un diseño ideado por John Tukey que proporciona una impresión visual rápida del número de observaciones o de datos de una clase. Cada observación del conjunto de datos se divide en dos partes: tallo y hoja, aunque hay bastante flexibilidad en cuanto al procedimiento que pueda seguirse, en ocasiones es conveniente considerar todos los dígitos de una observación menos el último como el tallo y éste último dígito se considera como la hoja.

Entre las ventajas que tiene podemos mencionar que es más fácil de construir que un histograma y dentro de un intervalo de clase, este gráfico de más información que un histograma porque muestra los valores reales.

Ejemplo:

Construir un diagrama de tallo y hoja para la colección de 25 calificaciones en un examen de álgebra:

78	59	81	65	80
64	65	79	85	54
98	82	75	59	89
67	57	68	71	67
84	87	65	76	94

Pasos para la construcción:

1. Coloque los tallos en forma vertical usando un segmento de línea vertical, llamado tronco, para separar los tallos de las hojas:

5	
6	
7	
8	
9	

2. Coloque cada hoja a la derecha de su tallo:

5	9 7 4 9
6	4 5 7 5 7 8 5
7	8 6 1 9 5
8	5 4 2 9 7 1 0
9	8 4

Aunque no importa el orden en que las hojas se coloquen en un tallo, es recomendable que se ordenen porque esto facilita el conteo.

3. Cabe resaltar que no hay una cantidad única de renglones o tallos, si creemos que nuestro diagrama original condensa demasiado los datos, podemos alargarlo usando dos o más renglones para cada uno o más dígitos.
   Ejemplo:

6	8 9
7	2 3 3
7	5 6 6
8	0 1 1 2 3 4
8	5 6
9	1 2 2 4

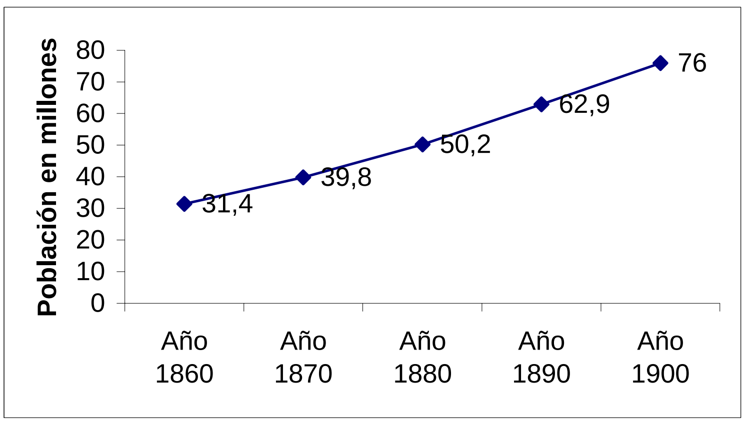
Es posible encontrar la siguiente notación:

5* | 9 7 4 9 6* | 4 5 7 5 6 7* | 4 5 6 8 9 0 El asterisco es para indicar que el número es de dos cifras. Ejemplo: 59,57,64, etc.	12* | 9 7 4 9 13* | 4 5 7 5 6 14* | 4 5 6 8 9 0 El asterisco es para indicar que el número es de tres cifras. Ejemplo: 129, 135, etc.	3** | 45 75 61 4** | 45 68 90 Los asteriscos son para indicar que el número es de tres cifras. Ejemplo: 345, 468, etc.

- **Gráfico de trazos**

Es un tipo de gráfico en donde se localizan los puntos en un sistema de coordenadas y luego se conectan los puntos sucesivos con trazos rectos.
Ejemplo:
La tabla muestra la población de USA (en millones) en los años de 1860 a 1900

Año	1860	1870	1880	1890	1900
Población	31,4	39,8	50,2	62,9	76,0



Se debe indicar el cero siempre que sea posible; en caso de que no lo sea, y si tal omisión pudiera provocar alguna conclusión errónea, es aconsejable advertirlo de algún modo (por lo general, con un corte en el eje).

- **Gráfico de barras**

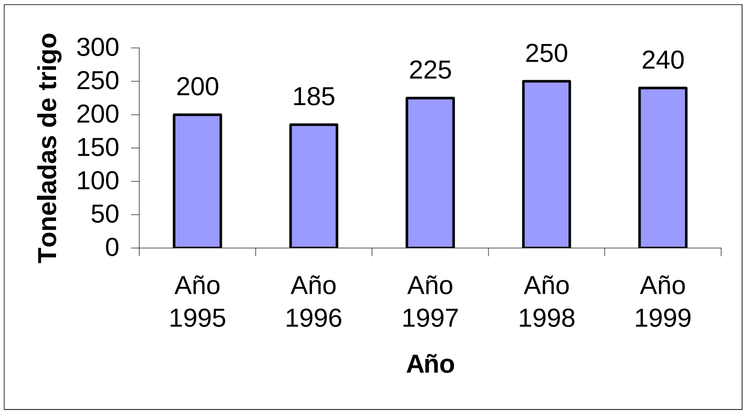
Consiste en una serie o conjunto de rectángulos que de acuerdo a su longitud y anchura representan un fenómeno, se puede utilizar para representar datos cualitativos y cuantitativos.
Observaciones:

1. En el eje donde irá la base del rectángulo se especifican los indicadores o nombres que se usan para cada una de las bases.
2. La escala que se debe tomar para la base debe ser la misma para cada rectángulo.
3. La separación que exista entre las barras debe ser la misma, depende del número de barras a construir y del espacio con que se cuenta.
4. En el eje vertical se puede representar una escala de frecuencias, frecuencias relativas o de porcentajes.

Entre los tipos de gráficos de barra tenemos:

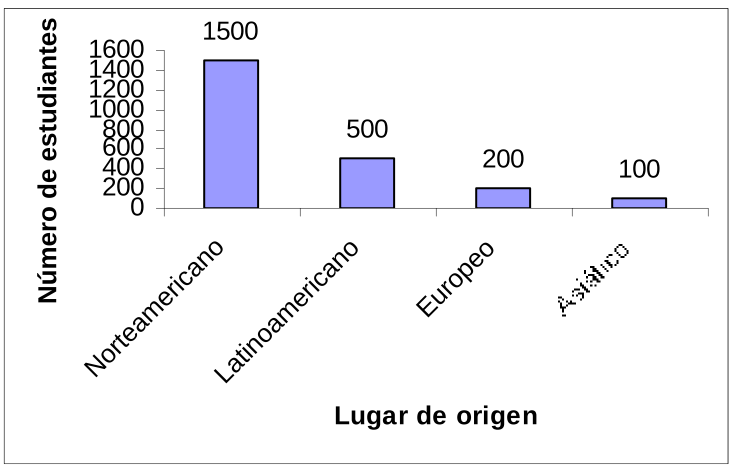
a. Gráficos de barras simples: son aquellos que representan una sola característica.
Ejemplo cuantitativo:
La tabla muestra el número de toneladas de trigo producidos por una cooperativa durante los años 1995 al 1999.

Año	1995	1996	1997	1998	1999
Toneladas de trigo	200	185	225	250	240



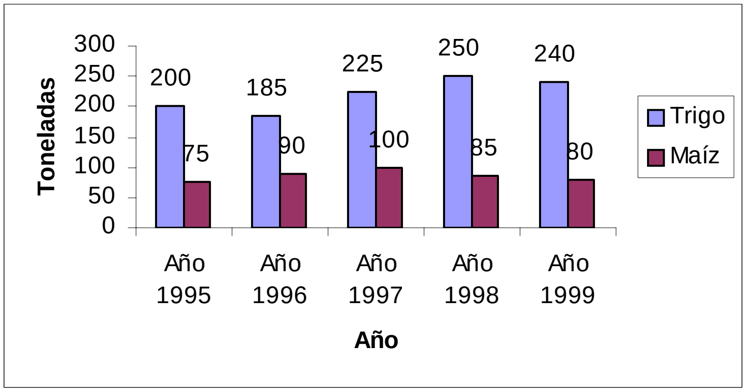
Ejemplo cualitativo:
Los siguientes datos corresponden al número de estudiantes de cierta universidad, de acuerdo con su lugar de origen.

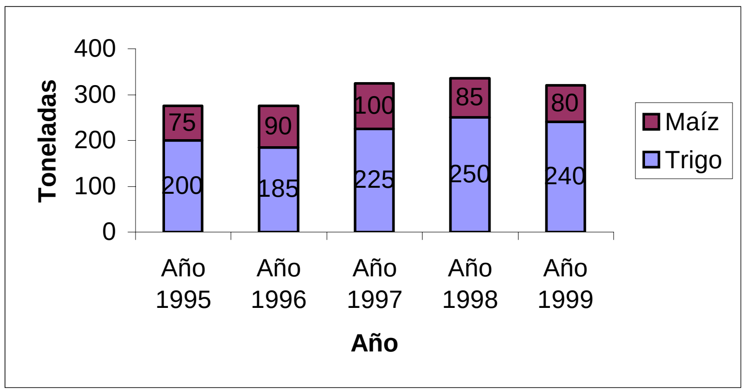
Lugar de origen	Norteamericano	Latinoamericano	Europeo	Asiático
Número de estudiantes	1500	500	200	100



b. Gráficos de barras compuestas: son aquellos que representan varias características, siendo útiles para propósitos comparativos.
Ejemplos cuantitativos:
La tabla muestra el número de toneladas de trigo y de maíz producidos por una cooperativa durante los años 1995 al 1999

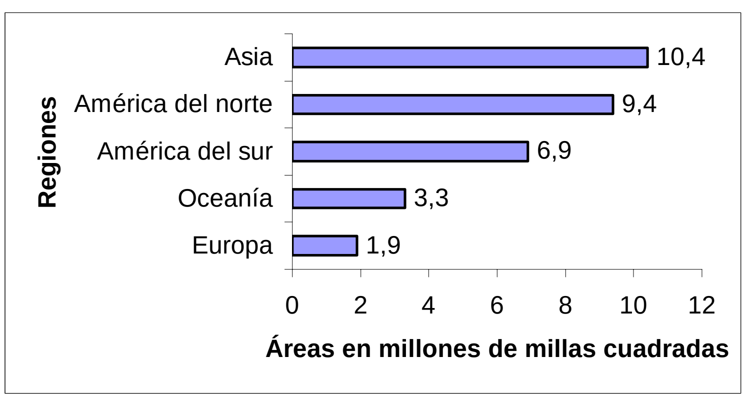
Año	1995	1996	1997	1998	1999
Toneladas de trigo	200	185	225	250	240
Toneladas de maíz	75	90	100	85	80





Es posible realizar los gráficos de barras no sólo en forma vertical, sino también horizontal.
Ejemplo:
Las áreas de algunas regiones (en millones de millas cuadradas) están dadas en la siguiente tabla:

Área	1,9	3,3	6,9	9,4	10,4
Región	Europa	Oceanía	América del sur	América del norte	Asia



❖ Histogramas

Son gráficos de barras en los cuales no hay separación entre los rectángulos que se forman, se construye mediante la representación de las clases de una distribución de frecuencias en el eje horizontal y las frecuencias en el eje vertical. A través de él se pueden visualizar tres características de los datos: forma, acumulación o tendencia posicional y la dispersión o variabilidad

Pasos para la construcción:

1. Se trazan dos ejes de coordenadas sobre un plano.
2. Se llevan sobre el eje horizontal a los límites de clase.
3. En el eje vertical podemos representar no sólo el número de frecuencias, también podemos colocar la proporción y el porcentaje de observaciones para cada intervalo de clase, por eso tenemos varios tipos de nombres:

Sobre el eje vertical	Nombre
Número de observaciones.	Histograma de frecuencias.
Proporción de observaciones.	Histograma de frecuencias relativas.
Porcentaje de observaciones.	Histograma porcentual

4. Se levantan perpendiculares por los límites de cada clase hasta la frecuencia de clase respectiva.
5. Se unen las dos perpendiculares que representan cada clase.

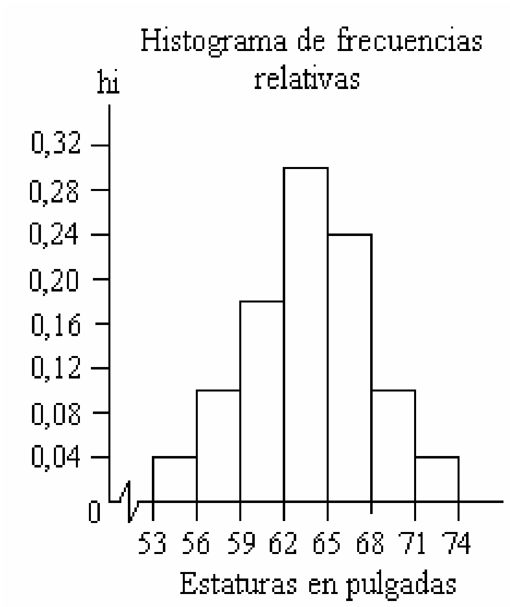
Observaciones:

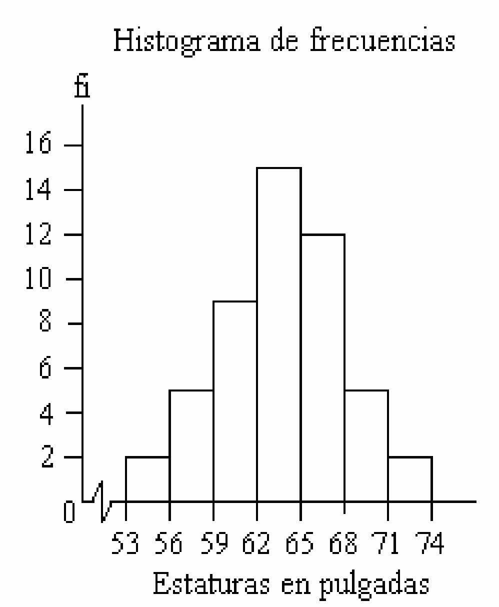
1. Los histogramas no se pueden utilizar con respecto a distribuciones de frecuencias de clases abiertas (a menos que la persona cierre el intervalo de una manera conveniente).
2. El histograma representa las frecuencias de los intervalos mediante áreas y no mediante alturas; sin embargo, si los intervalos de clase tienen todos igual tamaño entonces el área de los rectángulos representa las frecuencias, por ello las alturas de los rectángulos son proporcionales a las frecuencias de clase y se acostumbra en tal caso a tomar las alturas numéricamente iguales a las frecuencias de clase.
   Si los intervalos de clase no son de igual tamaño, las áreas no representan a las frecuencias, por lo tanto, es necesario ajustar la altura de los rectángulos (estas alturas deberán ser calculadas para que las superficies sean proporcionales a las frecuencias de clase).

Del ejemplo inicial sobre el investigador que deseaba determinar cómo variaban las estaturas de las obreras de una empresa y el cual tomaba una muestra de 50 mujeres para registrar luego sus estaturas en pulgadas, se obtuvo la siguiente tabla:

$l_i - l_{i+1}$	f_i	x_i	F_i	h_i	$\% h_i$	H_i	$\% H_i$
[53 – 56)	2	54,5	2	0,0400	4	0,0400	4
[56 – 59)	5	57,5	7	0,1000	10	0,1400	14
[59 – 62)	9	60,5	16	0,1800	18	0,3200	32
[62 – 65)	15	63,5	31	0,3000	30	0,6200	62
[65 – 68)	12	66,5	43	0,2400	24	0,8600	86
[68 – 71)	5	69,5	48	0,1000	10	0,9600	96
[71 – 74)	2	72,5	50	0,0400	4	1,0000	100

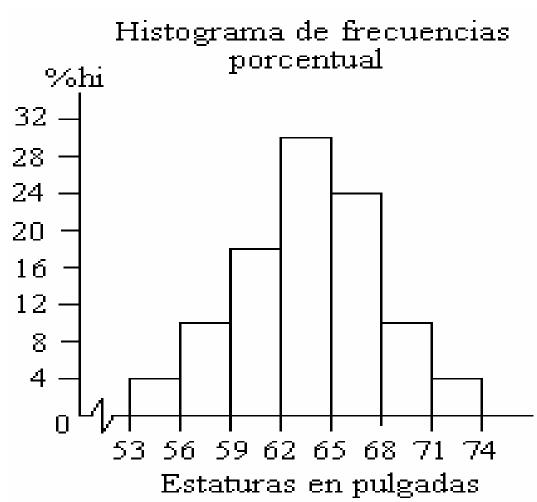
Utilizando los valores de la tabla tenemos las siguientes gráficas:





Como se ve, los histogramas tienen la misma forma. Esto se debe por que en las situaciones anteriores el tamaño relativo de cada rectángulo es la frecuencia de esa clase comparada con el número total de observaciones.

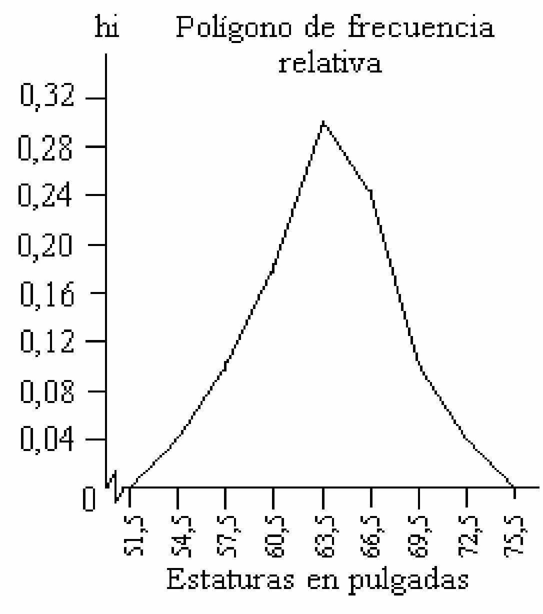
Por último:

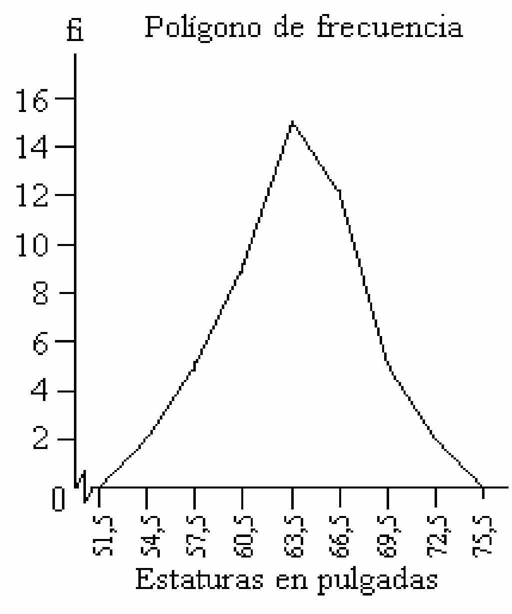


- **Polígonos de frecuencias**

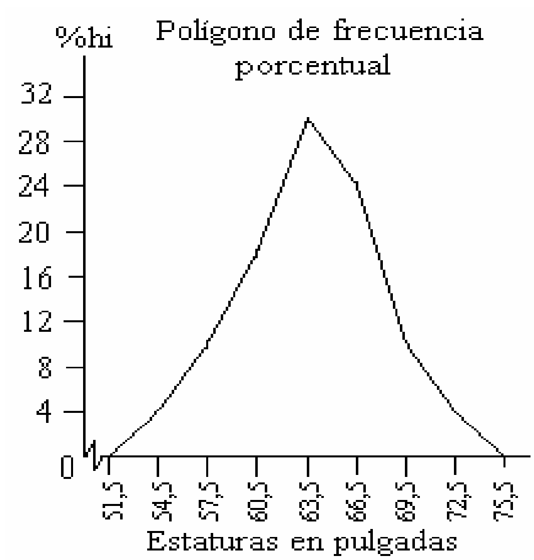
Son gráficos de línea trazados sobre las marcas de clase de cada intervalo, puede obtenerse uniendo los puntos medios de los techos de los rectángulos del histograma y tomando en cuenta que se deben extender ambos extremos del polígono hasta el eje horizontal en aquellos puntos que serían las marcas de clase adyacentes a cada extremo.

A medida que crece el número de clases y de observaciones, el polígono se vuelve cada vez más suave y curvo. Este polígono suavizado recibe el nombre de curva de frecuencia.
Al igual que sucede con los histogramas, tenemos el nombre del polígono según lo que se indique en el eje vertical; de esta forma tenemos polígonos de frecuencia, polígonos de frecuencia relativa y polígono porcentual.
Una de las ventajas de los polígonos es que nos permite hacer la comparación entre dos o más conjuntos de datos.



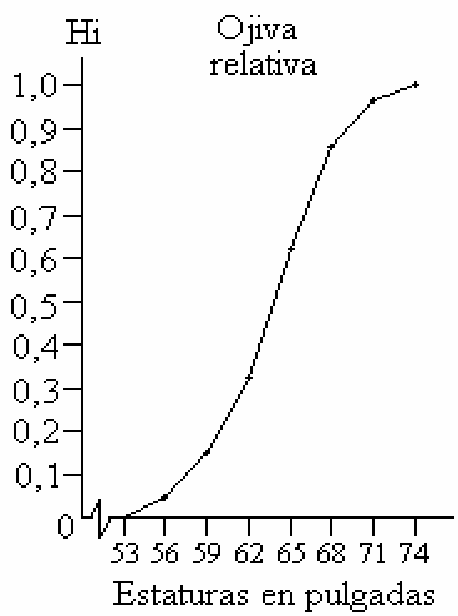


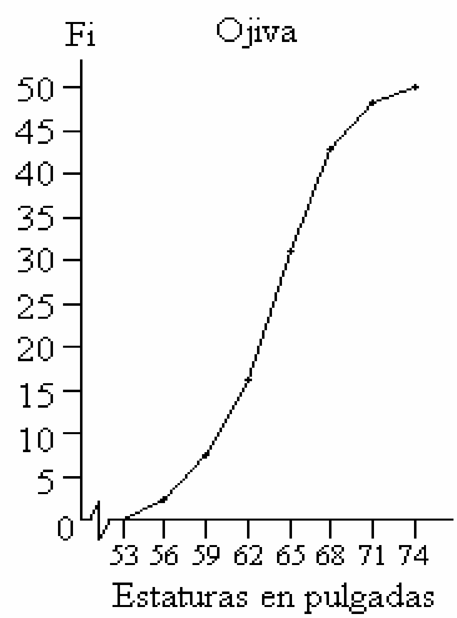
Por último:



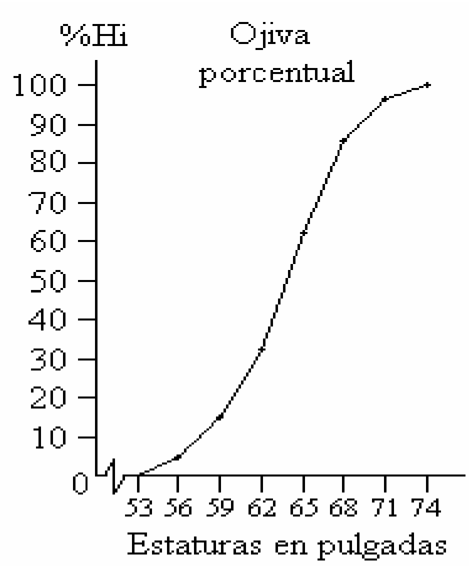
❖ Ojiva

Es la gráfica de una distribución de frecuencias acumuladas, los intervalos de las clases se ubican en el eje horizontal; las frecuencias acumuladas (ojiva propiamente dicha), las frecuencias relativas acumuladas (ojiva relativa) y las frecuencias acumuladas porcentuales (ojiva porcentual) se muestran en el eje vertical.
Podemos construir ojivas “o más” o las ojivas “menor que”, la diferencia entre ambas gráficas es que la primera tiene pendiente negativa y decrece, mientras que la segunda tiene pendiente positiva y crece. Se trabajarán ojivas del tipo “menor que” en este curso.
La ventaja de trabajar con ojivas es la facilidad (con respecto a otras gráficas) para interpolar entre los puntos trazados.
Tomando en cuenta los datos anteriores, las ojivas respectivas se presentan a continuación:





Por último:



Medidas descriptivas de las distribuciones de frecuencia.

Se ha visto que los métodos gráficos son extremadamente útiles para lograr una descripción de los datos y es por esto que las representaciones resultantes de las distribuciones de frecuencia nos permitieron discernir las tendencias y patrones de los datos; sin embargo, los métodos gráficos presentan limitaciones cuando se desea tener una mayor exactitud, motivo por el cual si necesitamos de medidas más exactas de un conjunto de datos, recurrimos a números individuales, llamados estadísticos resumidos. Mediante estos estadísticos podemos describir ciertas características del conjunto de datos los cuales nos permitirán tomar decisiones más rápidas y satisfactorias.

Cuatro de estas características son:

1) Medidas de tendencia central
2) Medidas de dispersión.
3) Medidas de sesgo.
4) Medidas de curtosis.

Medidas de tendencia central

Promedio

Es un valor típico o representativo de un conjunto de datos. Como tales valores tienden a situarse en el centro del conjunto de datos ordenados según su magnitud, los promedios se conocen también como medidas de centralización o de tendencia central.

Entre las medidas de tendencia central tenemos:

- **La Media Aritmética**

Es aquella que representa el promedio aritmético de un conjunto de observaciones, la misma actúa como punto de equilibrio, de manera que las observaciones menores equilibran a las mayores.

$$\text{Notación} \begin{cases} \bar{x} \text{ cuando sea para una muestra} \\ \mu \text{ cuando sea para una población} \end{cases}$$

Fórmulas:

Datos no agrupados	Datos agrupados
$\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} = \dfrac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$ $\mu = \dfrac{\sum_{i=1}^{N} x_i}{N}$	$\bar{x} = \dfrac{\sum_{i=1}^{g} x_i * f_i}{\sum f_i = n} = \dfrac{x_1 * f_1 + x_2 * f_2 + x_3 * f_3 + \ldots + x_g * f_g}{n}$ x_i son las marcas de clase f_i son las respectivas frecuencias absolutas

Para los datos agrupados, lo que se calcula es una estimación del valor de la media ya que al agrupar por clases no conocemos los valores individuales de cada observación, sólo que para facilitar los cálculos se ha de renunciar a la exactitud.

Ejemplos:

1) Calcular la media aritmética de 8, 3, 5, 12, 10:

$$\mu = \frac{8+3+5+12+10}{5} = 7{,}6$$

2) Del ejemplo inicial sobre el investigador que deseaba determinar cómo variaban las estaturas de las obreras de una empresa y el cual tomaba una muestra de 50 mujeres para registrar luego sus estaturas en pulgadas, se había obtenido la siguiente tabla:

l_i - l_{i+1}	f_i	x_i	F_i	h_i	% h_i	H_i	% H_i
[53 – 56)	2	54,5	2	0,0400	4	0,0400	4
[56 – 59)	5	57,5	7	0,1000	10	0,1400	14
[59 – 62)	9	60,5	16	0,1800	18	0,3200	32
[62 – 65)	15	63,5	31	0,3000	30	0,6200	62
[65 – 68)	12	66,5	43	0,2400	24	0,8600	86
[68 – 71)	5	69,5	48	0,1000	10	0,9600	96
[71 – 74)	2	72,5	50	0,0400	4	1,0000	100

Para calcular la media, debemos agregar una nueva columna:

l_i - l_{i+1}	f_i	x_i	x_i * f_i
[53 – 56)	2	54,5	109,0
[56 – 59)	5	57,5	287,5
[59 – 62)	9	60,5	544,5
[62 – 65)	15	63,5	952,5
[65 – 68)	12	66,5	798,0
[68 – 71)	5	69,5	347,5
[71 – 74)	2	72,5	145,0
			Σ = 3184

$$\bar{x} = \frac{3184}{50} = 63{,}68 \text{ pulgadas}$$

Interpretación: en promedio, las obreras presentaron una estatura de 63,68 pulgadas.

- **La Media Aritmética Ponderada**

A veces se asocia a los números de un conjunto de datos, ciertos factores o pesos y es por ello que la media aritmética ponderada es un promedio que se calcula a fin de tener en cuenta la importancia de cada valor para el total global.

Notación: $\bar{x}_w$

Fórmula:

$$\bar{x}_w = \frac{\sum_{i=1}^{k} w_i * x_i}{\sum_{i=1}^{k} w_i} = \frac{w_1 * x_1 + w_2 * x_2 + w_3 * x_3 + \ldots + w_k * x_k}{w_1 + w_2 + w_3 + \ldots + w_k}$$

Al calcular la media aritmética a partir de datos agrupados, en realidad obtuvimos la media aritmética ponderada utilizando las marcas de clase para los valores de “x” y las frecuencias de cada clase como los pesos, en ese caso $\Sigma f_i = \Sigma w_i$.

Ejemplo:
Si un examen final de curso se valora como 3 veces los exámenes parciales y un estudiante tiene una nota de examen final de 85 y notas de exámenes parciales de 70 y 90, calcular su nota final.

$\bar{x}_w = \dfrac{1*70 + 1*90 + 3*85}{1+1+3} = 83$ puntos.

Observaciones sobre la media aritmética:

- ✓ Es una medida que toma en consideración todos los valores de la distribución. Esto es positivo, pero por la misma razón es muy sensible a la presentación de observaciones extremas o anómalas que hacen que la media se desplace hacia ellas. En consecuencia no es recomendable usar la media como medida de tendencia central en los casos en el cual el conjunto de datos no es homogéneo, pues la cantidad obtenida no es representativa del total de los datos.
- ✓ Tiene la ventaja de que es única y siempre se puede calcular (si no hay intervalos abiertos).
- ✓ El valor de la media aritmética puede no coincidir con los valores de la variable.

Algunas propiedades de la media aritmética:

- ✓ La suma algebraica de las desviaciones de un conjunto de números con respecto a su media aritmética es cero.
  $$\sum_{i=1}^{k}\left(x_i - \bar{x}\right) = 0$$
- ✓ La media aritmética de una constante es igual a la constante.
- ✓ La media de la suma de dos o más variables es igual a la suma de las medias de dichas variables.
  $$\bar{x}\left(x_i + y_i\right) = \bar{x}\,x_i + \bar{x}\,y_i$$
- ✓ Si a cada valor de la serie se le agrega una constante, la media de la nueva serie es igual a la media de la serie original más la constante. Igual sucede si a la media se le resta una constante.
  $$\bar{x}\left(x_i + k\right) = \bar{x}\,x_i + k \qquad \bar{x}\left(x_i - k\right) = \bar{x}\,x_i - k$$
- ✓ Media de medias: Si f_1 números tienen de media m_1, f_2 números tiene de media m_2 ,..., f_k números tienen de media m_k, entonces la media de todos los números es:
  $$\bar{x}_w = \frac{f_1 * m_1 + f_2 * m_2 + f_3 * m_3 + \ldots + f_k * m_k}{f_1 + f_2 + f_3 + \ldots + f_k}$$
  es, decir, la media aritmética ponderada de todas las medias.

❖ La Mediana

Es el punto medio de un conjunto de datos representando el valor más central en dicho conjunto, por lo que deja por encima y por debajo la misma cantidad de datos (una vez que estos han sido ordenados). Geométricamente es el valor de “x” que corresponde a la vertical que divide al histograma en dos partes de igual área.
Notación: Med
Fórmulas:

Datos no agrupados	Datos agrupados
El valor de la mediana puede coincidir o no con un valor de la serie, todo depende si el número de datos es par o impar. Los pasos son: 1. Organizar por orden ascendente a los datos. 2. Utilizar la fórmula de posicionamiento de punto: $\frac{n+1}{2}$ para localizar el lugar que ocupa el valor de la mediana en el arreglo ordenado. 3. Si el conjunto tiene un número impar de elementos, el de la mitad será la mediana, si contiene un número par de elementos, la mediana será el promedio aritmético de los dos que se hallan en la mitad.	$Med = l_i + \frac{\frac{n}{2} - F_a}{f_{med}} * a$ en donde: l_i es el límite inferior. F_a es la frecuencia acumulada anterior. f_{med} es la frecuencia absoluta del intervalo de la mediana. a es la amplitud. Los pasos son: 1. Calcular $\frac{n}{2}$ 2. Localizar ese valor en F_i, si no está, pasar al inmediato superior. Con esto se halla el intervalo de la mediana. 3. Aplicar la fórmula sustituyendo los valores correspondientes.

Ejemplos:

1) Datos no agrupados:
   Sean los números: 3, 4, 4, 5, 6, 8, 8, 8, 10. Usando la fórmula de posicionamiento, el valor ocupado por la posición $\frac{9+1}{2} = 5$ sería a mediana, entonces la respuesta es 6.
   Sean los números: 5, 5, 7, 9, 11, 12, 15, 18. Usando la fórmula de posicionamiento, el valor $\frac{8+1}{2} = 4{,}5$ daría la posición de la mediana; como no hay esa posición, buscamos el promedio de los números que ocupan los puestos 4 y 5, dando como resultado que la mediana será 10

2) Datos agrupados:
   Del ejemplo inicial sobre el investigador que deseaba determinar cómo variaban las estaturas de las obreras de una empresa y el cual tomaba una muestra de 50 mujeres para registrar luego sus estaturas en pulgadas, se había obtenido la siguiente tabla:

l_i - l_{i+1}	f_i	x_i	F_i	h_i	% h_i	H_i	% H_i
[53 – 56)	2	54,5	2	0,0400	4	0,0400	4
[56 – 59)	5	57,5	7	0,1000	10	0,1400	14
[59 – 62)	9	60,5	16	0,1800	18	0,3200	32
[62 – 65)	15	63,5	31	0,3000	30	0,6200	62
[65 – 68)	12	66,5	43	0,2400	24	0,8600	86
[68 – 71)	5	69,5	48	0,1000	10	0,9600	96
[71 – 74)	2	72,5	50	0,0400	4	1,0000	100

Paso 1:

$$\frac{n}{2}=\frac{50}{2}=25$$

Paso 2:
Como 25 no aparece en F_i, pasamos al inmediato superior: 31.

Paso 3:

$$\text{Med}=62+\frac{25-16}{15}*3=63{,}8 \text{ pulgadas}$$

Interpretación: El 50% de las obreras tienen una estatura igual o inferior a 63,8 pulgadas aproximadamente.

Observaciones sobre la mediana:
- ✓ Como medida descriptiva, tiene la ventaja de no estar afectada por las observaciones extremas, ya que no depende de los valores que toma la variable, sino el orden de los mismos. Por ello, es adecuado su uso en distribuciones que presentan observaciones extremadamente grandes o pequeñas.
- ✓ Puede ser calculada aún a partir de datos agrupados con clases abiertas.
- ✓ Puede usarse con datos cualitativos.
- ✓ No utiliza toda la información de los datos (sólo los valores centrales).
- ✓ Su mayor defecto es que no se ajusta fácilmente al cálculo algebraico, lo que hace que sea difícil de utilizar en otras áreas, como en la inferencia.

❖ **La Moda**

Es el valor de los datos que se presenta con más frecuencia, por lo que representa el punto más alto en la curva de distribución de un conjunto de datos.
Notación: Mo
Fórmulas:

Datos no agrupados	Datos agrupados
No hay fórmulas, sólo ver cuál valor o elemento es el que más se repite.	$\text{Mo}=l_i+\frac{\Delta_1}{\Delta_1+\Delta_2}*a$ l_i es el límite inferior (si se trabajan con límites imaginarios y reales, se toman los reales). Δ_1 es el valor que se obtiene a restar la f_{modal} con la frecuencia anterior. Δ_2 es el valor que se obtiene a restar la f_{modal} con la frecuencia siguiente. a es la amplitud.

Los pasos para calcular la moda con datos agrupados serían:
1. Ubicar la mayor f_i para hallar el intervalo modal
2. Aplicar la fórmula

Del ejemplo inicial sobre el investigador que deseaba determinar cómo variaban las estaturas de las obreras de una empresa y el cual tomaba una muestra de 50 mujeres para registrar luego sus estaturas en pulgadas, se había obtenido la siguiente tabla:

l_i - l_{i+1}	f_i	x_i	F_i	h_i	% h_i	H_i	% H_i
[53 – 56)	2	54,5	2	0,0400	4	0,0400	4
[56 – 59)	5	57,5	7	0,1000	10	0,1400	14
[59 – 62)	9	60,5	16	0,1800	18	0,3200	32
[62 – 65)	15	63,5	31	0,3000	30	0,6200	62
[65 – 68)	12	66,5	43	0,2400	24	0,8600	86
[68 – 71)	5	69,5	48	0,1000	10	0,9600	96
[71 – 74)	2	72,5	50	0,0400	4	1,0000	100

Para calcular la moda:

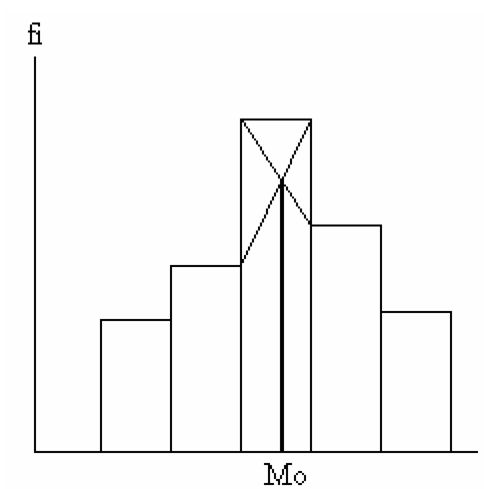
1. Ubicamos la mayor frecuencia absoluta, en este caso es 15 y el mismo pertenece a la cuarta clase.
2. Aplicamos la fórmula:

   $$\text{Mo} = 62 + \frac{15-9}{(15-9)+(15-12)} * 3 = 64 \text{ pulgadas}$$

   Interpretación: la mayoría de las obreras tienen una estatura de 64 pulgadas aproximadamente.

Observaciones sobre la moda:

- ✓ Se puede usar para datos cualitativos y cuantitativos.
- ✓ Se puede emplear aunque existan clases abiertas en la distribución.
- ✓ Puede no ser única, por ello, cuando los conjuntos de datos contiene 2, 3, o más modas, son difíciles de interpretar.
- ✓ Puede que una distribución no tenga moda.
- ✓ El intervalo modal es aquel que posee una barra en el histograma con mayor altura geométricamente, se calcula según la gráfica:



Cuantiles

Si una serie de datos se colocan en orden de magnitud, el valor medio que divide al conjunto de datos en dos partes iguales es la mediana, por extensión de esta idea se puede pensar en aquellos valores que dividen a los datos en cuatro partes iguales, en cien partes iguales, etc. El nombre genérico es el de cuantil y el mismo se define como el valor bajo el cual se encuentra una determinada proporción de los valores de una distribución.
Dentro de las medidas de los cuantiles tenemos:

- Deciles:

Son aquellos valores que dividen en diez partes iguales a un conjunto de datos ordenados. Se representan por D_1 , D_2 , D_3 , ...D_9. De esta manera tenemos que:
- D_1 (primer decil) es el valor por debajo del cual se encuentran como máximo el 10% de las observaciones, mientras que el 90% restante se sitúan por encima de él.
- D_2 (segundo decil) es el valor por debajo del cual se encuentran como máximo el 20% de las observaciones, mientras que el 80% restante se sitúan por encima de él.

Y así sucesivamente.

- Cuartiles:

Son aquellos valores que dividen en cuatro partes iguales a un conjunto de datos ordenados. Se representan por Q_1, Q_2, y Q_3. De esta manera tenemos que:
- Q_1 (primer cuartil) es el valor por debajo del cual se sitúan a lo sumo el 25% de las observaciones y por encima de éste el 75% restante.
- Q_2 (segundo cuartil) es el valor por debajo de cual se sitúan a lo sumo el 50% de las observaciones y por encima de éste el 50% restante. Está justo en el centro y corresponde a la mediana
- Q_3 (tercer cuartil) es el valor por debajo del cual se sitúan a lo sumo el 75% de las observaciones y por encima de éste el 25% restante

Observación: Hay algunas variaciones en las convenciones de cálculo de cuartiles ya que los valores reales calculados pueden variar un poco dependiendo de la convención seguida. Sin embargo, el objetivo de todos los procedimientos de cálculo de cuartiles es dividir los datos en aproximadamente cuatro partes iguales.

- Percentiles:

Son aquellos valores que dividen a un conjunto de datos ordenados en cien partes iguales. Se representan por P_1, P_2 .... , P_{99}. De esta manera tenemos que:
- P_1 es el valor por debajo del cual se sitúan a lo sumo el 1% de los datos y por encima de él tenemos el 99% restante.
- P_2 es el valor por debajo del cual se sitúan a lo sumo el 2% de los datos y por encima de él tenemos el 98% restante. Y así sucesivamente..

En forma genérica el p-ésimo percentil es un valor tal que por lo menos un “p” por ciento de los elementos tiene dicho valor o menos y, al menos, un (100-p) por ciento de los elementos tiene ese valor o más.
Es conveniente tomar en cuenta que: $D_1 = P_{10}$, $D_2 = P_{20}$, $Q_1 = P_{25}$, y así sucesivamente.
Dependiendo de si trabajamos con datos agrupados o no tendremos los siguientes pasos para el cálculo de los percentiles:

* Para datos no agrupados:
    1. Ordenar los datos de manera ascendente.
    2. Calcular el índice:

$$i = \left(\frac{P}{100}\right) * n$$

- Si "i" es entero, el p-ésimo percentil es el promedio de los valores de los datos ubicados en los lugares "i" e "i +1".

- Si "i" no es entero, se redondea. El valor entero inmediato mayor que "i" indica la posición del p-ésimo percentil.

Ejemplo:
Determinar el P_{50} y el P_{85} de los datos siguientes:
2350, 2450, 2550, 2380, 2255, 2210, 2390, 2630, 2440, 2825, 2420, 2380.
1. Ordenamos de manera ascendente:
   2210, 2255, 2350, 2380, 2380, 2390, 2420, 2440, 2450, 2550, 2630, 2825.

Para P_{50}:

2. Calculamos "i": $i = \left(\frac{50}{100}\right) * 12 = 6$

   Como "i" es entero, P_{50} es el promedio de los 6° y 7°, luego nos daría:

$$P_{50} = \frac{2390 + 2420}{2} = 2405$$

Para P_{85}:

2. Calculamos "i": $i = \left(\frac{85}{100}\right) * 12 = 10{,}2$

Como "i" no es entero, redondeamos. El lugar del P_{85} es el siguiente entero mayor que 10,2 es decir, el lugar 11. Esto nos daría que $P_{85} = 2630$.

* Para datos agrupados:
    a. Se aplica la fórmula:

$$P_p = l_i + \frac{\frac{n * p}{100} - F_a}{f_p} * a$$

b. Para aplicar la fórmula, los pasos son:
   1. Ubicar el resultado de $\frac{n * p}{100}$ en F_i
   2. Si no está el valor, se pasa al inmediato superior.
   3. Al ubicar el valor de F_i determinamos el intervalo de donde se obtendrán los datos para sustituir en la ecuación.
      Observación: si se trabajan con límites reales e imaginarios, se toman los reales.

Por medio de los percentiles, se halla el valor de la variable para un porcentaje dado.

Del ejemplo inicial sobre el investigador que deseaba determinar cómo variaban las estaturas de las obreras de una empresa y el cual tomaba una muestra de 50 mujeres para registrar luego sus estaturas en pulgadas, se había obtenido la siguiente tabla:

l_i - l_{i+1}	f_i	x_i	F_i	h_i	% h_i	H_i	% H_i
[53 – 56)	2	54,5	2	0,0400	4	0,0400	4
[56 – 59)	5	57,5	7	0,1000	10	0,1400	14
[59 – 62)	9	60,5	16	0,1800	18	0,3200	32
[62 – 65)	15	63,5	31	0,3000	30	0,6200	62
[65 – 68)	12	66,5	43	0,2400	24	0,8600	86
[68 – 71)	5	69,5	48	0,1000	10	0,9600	96
[71 – 74)	2	72,5	50	0,0400	4	1,0000	100

$P_{15} = ?$

1. $\frac{n*p}{100} = \frac{50*15}{100} = 7{,}5$
2. $P_{15} = 59 + \frac{7{,}5-7}{9} * 3 = 59{,}17$ pulgadas

Interpretación: el 15% de las obreras tienen una estatura de 59,17 pulgadas o menos.

$Q_1 = ?$
$Q_1 = P_{25}$

1. $\frac{n*p}{100} = \frac{50*25}{100} = 12{,}5$
2. $P_{25} = 59 + \frac{12{,}5-7}{9} * 3 = 60{,}83$ pulgadas

Interpretación: el 25% de las obreras tienen una estatura de 60,83 pulgadas o menos.

$D_3 = ?$
$D_3 = P_{30}$

1. $\frac{n*p}{100} = \frac{50*30}{100} = 15$
2. $P_{30} = 59 + \frac{15-7}{9} * 3 = 61{,}67$ pulgadas

Interpretación: el 30% de las obreras tienen una estatura de 61,67 pulgadas o menos.

$P_{75} = ?$

1. $\frac{n*p}{100} = \frac{50*75}{100} = 37{,}5$
2. $P_{75} = 6 + \frac{37{,}5-31}{12} * 3 = 66{,}63$ pulgadas

Interpretación: el 75% de las obreras tienen una estatura de 66,63 pulgadas o menos.

❖ **Rango Percentil**

Es una expresión mediante la cual podemos hallar el porcentaje, dado un valor de la variable. Dicha expresión se obtiene al despejar "p" en la fórmula de percentiles para datos agrupados, el proceso para hallar el rango percentil es:

1. Ubicar el valor de la variable que nos dan, en el intervalo que le corresponda.
2. Una vez ubicado, podemos determinar l_i, f_i, etc, para sustituir en la fórmula:

$$p = \frac{\frac{(P_p - l_i) * f_i}{a} + F_a}{n} * 100$$

Del ejemplo inicial sobre el investigador que deseaba determinar cómo variaban las estaturas de las obreras de una empresa y el cual tomaba una muestra de 50 mujeres para registrar luego sus estaturas en pulgadas, se había obtenido la siguiente tabla:

l_i - l_{i+1}	f_i	x_i	F_i	h_i	% h_i	H_i	% H_i
[53 – 56)	2	54,5	2	0,0400	4	0,0400	4
[56 – 59)	5	57,5	7	0,1000	10	0,1400	14
[59 – 62)	9	60,5	16	0,1800	18	0,3200	32
[62 – 65)	15	63,5	31	0,3000	30	0,6200	62
[65 – 68)	12	66,5	43	0,2400	24	0,8600	86
[68 – 71)	5	69,5	48	0,1000	10	0,9600	96
[71 – 74)	2	72,5	50	0,0400	4	1,0000	100

Hallar el porcentaje de obreras cuyas estaturas son iguales o inferiores a 67 pulgadas:

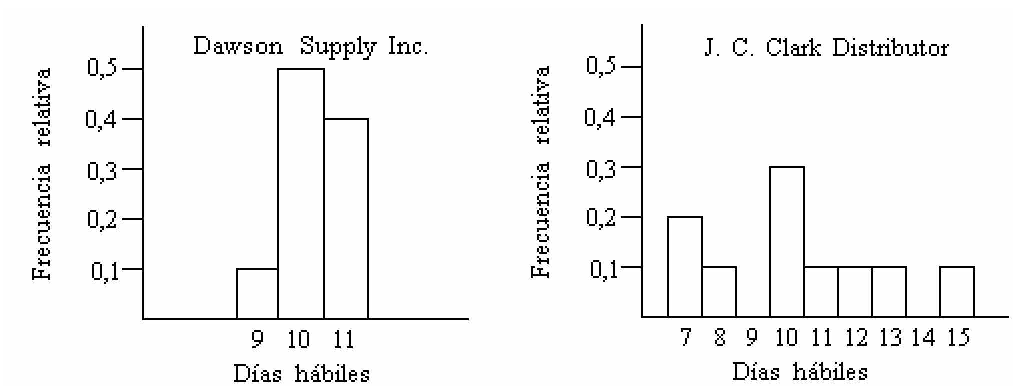
1. Ubicamos el valor de 67 en la tabla y vemos que corresponde a la 5[ta] clase.
2. Sustituimos los valores:

$$p = \frac{\frac{(67 - 65) * 12}{3} + 31}{50} * 100 = 78\%$$

Interpretación: el 78% de las obreras tienen estaturas iguales o inferiores a 67 pulgadas.

Medidas de dispersión

Mientras los estadísticos de tendencia central nos indican los valores alrededor de los cuales se sitúan un grupo de observaciones, los estadísticos de variabilidad o dispersión muestran si los valores de las observaciones están próximos entre sí o están muy separados.
Dos conjuntos de datos pueden tener la misma localización central y no obstante, ser muy distintos si uno se halla más disperso que el otro.
Por ejemplo, supongamos que usted es un agente de compras de una importante empresa manufacturera, y con regularidad coloca pedidos con dos proveedores distintos. Ambos le indican que necesitan alrededor de 10 días hábiles para surtir sus pedidos. Después de varios meses de trabajar así encuentra usted que el promedio de días necesarios para surtir los pedidos es, realmente, unos 10 para cada proveedor. Los histogramas que resumen la cantidad de días hábiles requeridos para surtir los pedidos se ven en la figura. Aunque la cantidad promedio es, más o menos, de 10 en ambos casos. ¿Tienen éstos el mismo grado de confiabilidad para entregar a tiempo?. Observe la dispersión, o variabilidad, en los histograma. ¿Qué proveedor prefiere usted?



Para la mayoría de las empresas es importante recibir materiales y suministros tiempo. Las entregas a los siete u ocho días de J. C. Clark Distributor pueden considerarse favorables; sin embargo, algunas de las entregas a los 13 o 15 días podrían ser desastrosas en términos de la utilización de la mano de obra y del cumplimiento de los programas de producción. Este ejemplo ilustra un caso en el que la dispersión, o variedad, en los tiempos de entrega puede ser la consideración más importante para seleccionar un proveedor. Para la mayoría de los agentes de compra, la menor dispersión que muestra Dawson Supply, Inc. haría que fuera el proveedor más consistente y preferido.

- **Dispersión**:

Es el grado en que los datos numéricos tienden a extenderse alrededor de un valor medio. La dispersión de la distribución suministra información complementaria que permite juzgar la confiabilidad de nuestra medida de tendencia central. Si los datos están ampliamente dispersos, la localización central será menos representativa de los datos en su conjunto de lo que sería en el caso de datos que se acumulasen más alrededor de la media. Además, si no conviene tener una amplia dispersión de valores respecto al centro o si esa dispersión implica un riesgo inaceptable, deberemos ser capaces de reconocerlo y no escoger las distribuciones que presentan la máxima dispersión.

Por ejemplo, a los analistas financieros les interesa la dispersión de las ganancias de una empresa, las utilidades con una fuerte dispersión indican un riesgo mayor parar los accionistas que las utilidades que permanecen relativamente estables.
Las medidas de dispersión se dividen en dos grandes grupos:

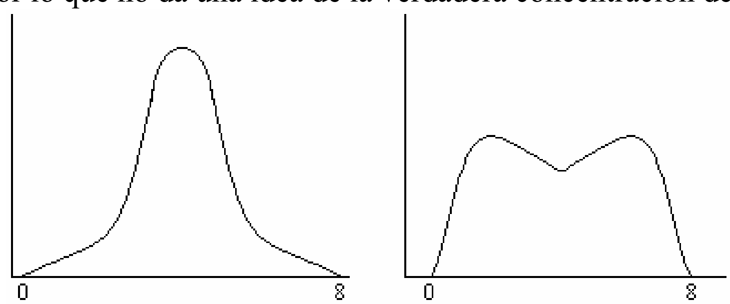
1- Las medidas de dispersión absolutas: son aquellas que vienen expresadas en las mismas medidas que identifican a la serie de datos.
2- Las medidas de dispersión relativas: son relaciones entre medidas de dispersión absolutas y medidas de tendencia central.

Medidas de dispersión absoluta:

❖ **Rango o recorrido**

Es la diferencia entre el valor más alto y el más bajo observado
Notación: R
Fórmula: R = $x_{máx} - x_{min}$
Un rango pequeño indica poca variación, uno grande indica una gran variabilidad.
Observaciones:

✓ No es muy útil porque sólo toma en cuenta los valores máximo y mínimo de una distribución por lo que no da una idea de la verdadera concentración de los valores.

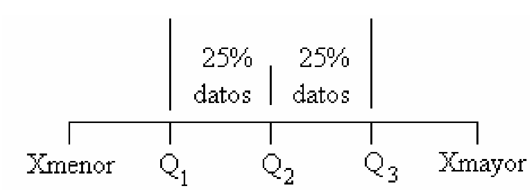


Igual rango, pero diferente variabilidad.

✓ No se puede utilizar en distribuciones que tengan intervalos abiertos.
✓ Puede ser afectado por observaciones externas.

❖ **Rango intercuartílico o rango intercuartil.**

Es la diferencia entre los valores de Q_1 y Q_3, esta diferencia refleja la variabilidad de las observaciones del 50% intermedio de los datos y tiene la ventaja de no verse influenciado por valores extremos.
Notación : RI
Fórmula: RI = $Q_3 - Q_1$
Gráficamente:



A través del rango intercuartil podemos ver (aproximadamente) qué tan lejos de la mediana tenemos que ir en cualquiera de las dos direcciones antes de que podamos recorrer una mitad de los valores del conjunto de datos.

Para los efectos de la situación que se ha mantenido como ejemplo, el rango intercuartil es:
Q_3-Q_1 = (66,63 – 60,83) pulgadas = 5,80 pulgadas.
Sin embargo, una medida más usada es el rango semi-intercuartil.

- **Rango semi-intercuartílico o rango semi-intercuartil.**

Es la semidiferencia entre los valores de Q_1 y Q_3, al igual que el rango intercuartílico tiene la ventaja de no verse influenciado por valores extremos.
Notación : RSI
Fórmula: $RSI = \dfrac{Q_3 - Q_1}{2}$

Utilizando el resultado anterior, tenemos que RSI = $\dfrac{66{,}63 - 60{,}83}{2} = 2{,}9$ pulgadas. De esto, pudiéramos establecer que un 50% de las estaturas caen en el intervalo [63,8±2,9] pulgadas.

- **Varianza**

Es la medida del cuadrado de la distancia promedio entre la media y cada elemento de la población.

Notación: $\begin{cases} \sigma^2 & \text{para la población} \\ s^2 & \text{para la muestra} \end{cases}$

Fórmulas:

Datos no agrupados	Datos agrupados
$\sigma^2 = \dfrac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$	$\sigma^2 = \dfrac{\sum_{i=1}^{g}(x_i - \mu)^2 * f_i}{N}$
$s^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$	$s^2 = \dfrac{\sum_{i=1}^{g}(x_i - \bar{x})^2 * f_i}{n-1}$

Nota: La teoría matemática establece que si pretendemos estimar la varianza de una población a partir de la varianza una de sus muestras, resulta que el error cometido es generalmente menor, cuando para la varianza de la muestra se divide por n –1 y no por n, porque el valor resultante da una mejor estimación de la varianza de la población.
Sin embargo, para grandes valores de n (n >30) no hay prácticamente diferencia entre dividir por n o por n-1.

Del ejemplo inicial sobre el investigador que deseaba determinar cómo variaban las estaturas de las obreras de una empresa y el cual tomaba una muestra de 50 mujeres para registrar luego sus estaturas en pulgadas, se había obtenido la siguiente tabla:

l_i - l_{i+1}	f_i	x_i	F_i	h_i	% h_i	H_i	% H_i
[53 – 56)	2	54,5	2	0,0400	4	0,0400	4
[56 – 59)	5	57,5	7	0,1000	10	0,1400	14
[59 – 62)	9	60,5	16	0,1800	18	0,3200	32
[62 – 65)	15	63,5	31	0,3000	30	0,6200	62
[65 – 68)	12	66,5	43	0,2400	24	0,8600	86
[68 – 71)	5	69,5	48	0,1000	10	0,9600	96
[71 – 74)	2	72,5	50	0,0400	4	1,0000	100

Para calcular la varianza agregamos una nueva columna:

l_i - l_{i+1}	f_i	x_i	$(x_i - \bar{x})^2 * f_i$
[53 – 56)	2	54,5	168,5448
[56 – 59)	5	57,5	190,9620
[59 – 62)	9	60,5	91,0116
[62 – 65)	15	63,5	0,4860
[65 – 68)	12	66,5	95,4288
[68 – 71)	5	69,5	169,3620
[71 – 74)	2	72,5	155,5848
			Σ = 871,38

$$s^2 = \frac{871{,}38}{49} = 17{,}7833 \text{ pulgadas}^2$$

Algunas propiedades de la varianza:

*La varianza de una constante es cero.
*Siempre es una cantidad positiva.
*La varianza del producto de una constante por una variable es igual al producto de la constante al cuadrado por la varianza de la variable.

Observaciones sobre la varianza:

- ✓ Las unidades de la varianza son los cuadrados de las unidades de los datos y en muchas ocasiones no son fáciles de interpretar.
- ✓ Puede sufrir un cambio desproporcionado por la existencia de valores extremos en el conjunto.

❖ Desviación típica o estándar.

Se define como la raíz cuadrada positiva de la varianza

$$\text{Notación: } \begin{cases} \sigma \text{ para la población} \\ s \text{ para la muestra} \end{cases}$$

Fórmulas:

Datos no agrupados	Datos agrupados
$\sigma = \sqrt{\dfrac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$	$\sigma = \sqrt{\dfrac{\sum_{i=1}^{g}(x_i - \mu)^2 * f_i}{N}}$
$s = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1}}$	$s = \sqrt{\dfrac{\sum_{i=1}^{g}(x_i - \bar{x})^2 * f_i}{n - 1}}$

Tomando el resultado de la varianza calculada anteriormente, s = 4,2170 pulgadas.

Algunas propiedades de la desviación típica:
*La desviación típica de una constante es cero.
*Siempre es una cantidad positiva.
*La desviación típica del producto de una constante por una variable es igual al producto de la constante por la desviación típica de la variable.

Observaciones sobre la desviación típica:

- ✓ Entre sus aplicaciones tenemos el teorema de Chebyshev, el cual afirma que para cualquier conjunto de datos, al menos $1 - \frac{1}{k^2}$ de la observaciones están dentro de k desviaciones típicas de la media (K >1). En virtud de esto, si por ejemplo, k = 2 nos daría 0,75. Lo que significa que si formamos un intervalo de 2 desviaciones típicas por debajo de la media hasta 2 desviaciones típicas por encima de la media, en dicho intervalo se encontrarán como mínimo el 75% de todas las observaciones.
- ✓ Nos permite determinar con mayor grado de precisión dónde se sitúan los valores de una distribución de frecuencia en relación con la media.
- ✓ Las unidades de la desviación típica se expresan en las mismas unidades de los datos.
- ✓ Puede sufrir un cambio desproporcionado por la existencia de valores extremos en el conjunto.

Medidas de dispersión relativa:

$$\text{Dispersión relativa} = \frac{\text{dispersión absoluta}}{\text{promedio}}$$

Estas medidas vienen generalmente expresadas en porcentajes y su función es la de determinar entre varias distribuciones la de mayor o menor dispersión, esto tiene como ventaja que nos permite comparar distribuciones donde las unidades pueden ser diferentes ya que estas medidas son independientes de las unidades utilizadas. Además, varias distribuciones pueden tener un mismo valor para determinada medida de dispersión y ser la variabilidad de sus datos en relación con la media, diferente
Se trabajará con:

❖ Coeficiente de variación.

Mide el grado de dispersión de un conjunto de datos en relación con su media.
Notación: CV
Fórmulas:

$$CV = \frac{s}{\bar{x}} * 100\% \quad \text{para la muestra}$$

$$CV = \frac{\sigma}{\mu} * 100\% \quad \text{para la población}$$

Observaciones:

- ✓ El CV es un estadístico útil para comparar la dispersión de conjuntos de datos que tienen distintas desviaciones estándar y distintos promedios.
- ✓ El CV pierde su utilidad cuando la media se aproxima a cero.

Para los efectos de la situación que se ha mantenido como ejemplo, tenemos
s = 4,2170 pulgadas.
$\bar{x} = 63{,}68$ pulgadas.

$$CV = \frac{4{,}22}{63{,}68} * 100\% = 6{,}62\%$$

Interpretación: la desviación típica de la muestra es el 6,62% del valor de la media de la muestra.

Es importante destacar que las medidas de dispersión relativa sirven para comparar las variabilidades de dos conjuntos de valores (poblaciones o muestras), mientras que si deseamos comparar a dos individuos de cada uno de esos conjuntos, es mejor usar valores tipificados.

Variables tipificadas

Los distintos conjuntos de datos están asociados por lo general a diferentes medias, ya sea porque son de naturaleza diferente o porque al ser la misma característica medida, sus centros no son los mismos. Con el propósito de reducir los datos a un mismo punto de referencia y a una escala común, se realiza entre ellos una transformación llamada tipificación.

Se conoce por tipificación de una variable "x" a efectuar el cambio de origen y de escala de la variable.
Notación: z

$$\text{Fórmulas: } \begin{cases} z = \dfrac{x - \bar{x}}{s} & \text{para muestras} \\ z = \dfrac{x - \mu}{\sigma} & \text{para población} \end{cases}$$

Esta nueva variable carece de unidades de medida y permite comparar dos o más cantidades que en un principio no son comparables porque aluden a conceptos diferentes. También es aplicable a casos en que se quieran comparar individuos semejantes de poblaciones diferentes. Por ejemplo, si deseamos comparar el nivel académico de dos estudiantes de diferentes universidades, z nos indica cuántas desviaciones estándar está un valor por arriba o por debajo de la media del conjunto de datos al cual pertenece.

Ejemplo:
Un estudiante obtuvo 84 puntos en el examen final de matemáticas, en el que la nota media fue 76, y la desviación típica 10. En el examen final de física obtuvo 90 puntos, siendo la media 82 y la desviación típica 16. ¿En qué examen sobresalió más?.

Examen de matemática

$\bar{x} = 76$
$s = 10$
$x = 84$
$z = \frac{84 - 76}{10} = 0,8$

Examen de física

$\bar{x} = 82$
$s = 16$
$x = 90$
$z = \frac{90 - 82}{16} = 0,5$

Sobresalió más en matemáticas.

Medidas de Sesgo y Curtosis

❖ **Medidas de sesgo**

En un análisis estadístico de una serie de valores, no sólo interesa conocer el promedio y la dispersión de los datos, sino también cómo se refleja o se acerca esta serie a una distribución simétrica.

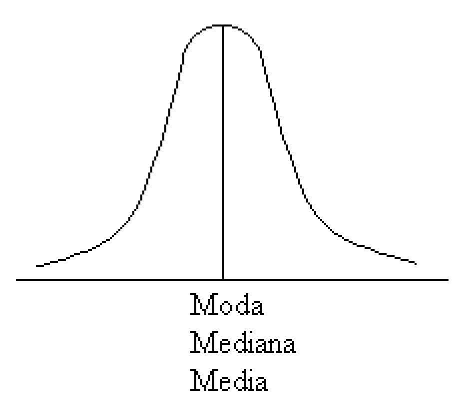
➢ **Sesgo**

Es el grado de asimetría de una distribución.

➢ **Curvas simétricas.**

Son aquellas en las cuales al trazar una línea vertical desde la cumbre de la curva al eje horizontal, se divide su área en dos partes iguales.

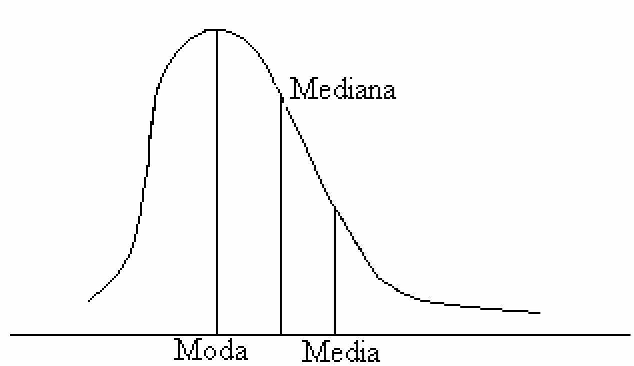
Gráficamente



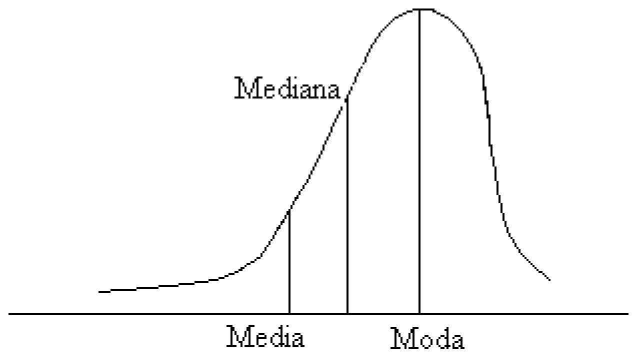
➢ **Curvas asimétricas**

Son aquellas curvas en las cuales al trazar una línea vertical desde su cumbre al eje horizontal, no se divide su área en dos partes iguales y pueden ser:

1) Asimetría positiva (sesgo a la derecha): es una curva que disminuye gradualmente hacia el extremo superior de la escala.



2) Asimetría negativa (sesgo a la izquierda): es una curva que disminuye gradualmente hacia el extremo inferior de la escala.



- **Coeficiente de asimetría de Pearson.**

Notación: SK
Fórmulas:

1. $SK = \dfrac{\bar{x} - Mo}{s}$
2. $SK = \dfrac{3\left(\bar{x} - Med\right)}{s}$

Si SK > 0 La asimetría es positiva.
Si SK = 0 Hay simetría.
Si SK < 0 La asimetría es negativa.

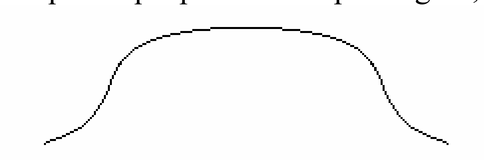
Del ejemplo inicial sobre el investigador que deseaba determinar cómo variaban las estaturas de las obreras de una empresa y el cual tomaba una muestra de 50 mujeres para registrar luego sus estaturas en pulgadas, se habían obtenido los siguientes valores:
$\bar{x} = 63{,}68$ pulgadas
Mo = 64 pulgadas
s = 4,2170 pulgadas
SK = -0,0759 asimetría negativa, sesgo a la izquierda.

- **Medidas de curtosis**

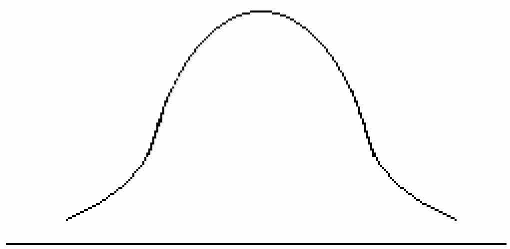
- **Curtosis**

Es el grado de pico o de apuntamiento que presenta una distribución. El patrón de referencia es la distribución normal o gaussiana.

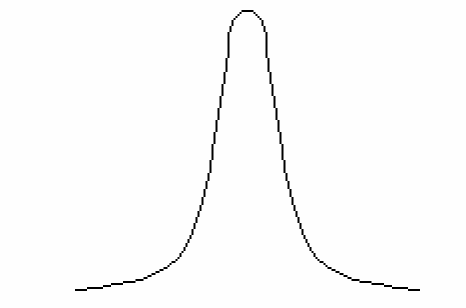
1) Curva platicúrtica: es aquella que presenta un pico ligero, es achatada.



2) Curva mesocúrtica: es aquella no es ni muy puntiaguda ni muy achatada (es la curva normal).



3) Curva leptocúrtica: es aquella que presenta un pico alto.



- **El coeficiente de curtosis.**

Es al medida que nos da una idea acerca del achatamiento o levantamiento de la curva en relación con la normal.
Notación: K

Para determinar la curtosis, se establece el porcentaje de valores que se encuentran en el intervalo $\bar{x} \pm s$ para considerar lo siguiente:
Si el resultado es menor a 68%, es platicúrica
Si el resultado es aproximadamente igual a 68%, es mesocúrtica
Si el resultado es mayor a 68%, es leptocúrtica

Del ejemplo inicial sobre el investigador que deseaba determinar cómo variaban las estaturas de las obreras de una empresa y el cual tomaba una muestra de 50 mujeres para registrar luego sus estaturas en pulgadas, se había obtenido: $\bar{x} = 63{,}68$ pulgadas y s = 4,22 pulgadas. Al calcular el porcentaje para el intervalo $\bar{x} \pm s$ se obtuvo 68,44%, por lo que es aproximadamente mesocúrtica.

Diagramas de bloques y líneas o boxplot

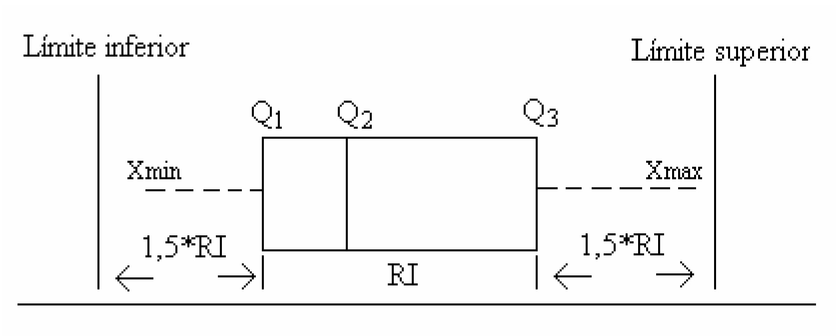
En su forma más simple, el diagrama de bloques y líneas ofrece una representación gráfica de los datos a través de los cinco números de resumen: X_{menor}, Q_1, Q_2, Q_3, X_{mayor}.

Pasos para construir un boxplot:

1) Construya una recta y marque en ella los 3 cuartiles.

2) Dibuje una caja sobre la recta con los extremos localizados en Q_1 y Q_3.
3) Trace un segmento vertical por el punto correspondiente a la mediana dentro de la caja, así la línea de la mediana divide los datos en 2 partes iguales.
4) Se ubican los límites mediante el rango intercuartil: los límites están a 1,5*RI debajo de Q_1 y a 1,5*RI arriba de Q_3. Se considera que los datos fuera de estos límites son valores atípicos.
5) Se trazan dos líneas punteadas (extensiones o bigotes de la caja): una que va del centro de la primera vertical hasta el valor mínimo dentro de los límites, y la otra que va del centro de la segunda vertical hasta el valor máximo dentro de los límites.
6) Se marcan con un asterisco las localizaciones de los valores atípicos.

El lugar ocupado por la mediana dentro de la caja es un buen indicador de la simetría, así, mirando la caja, si la línea trazada por la mediana está en el centro la distribución de los datos entonces tiende a ser simétrica, si la línea mediana se acerca al límite inferior, hay indicios de asimetría positiva y si está cerca del límite superior hay indicios de asimetría negativa. Gráficamente:



Ejemplo:

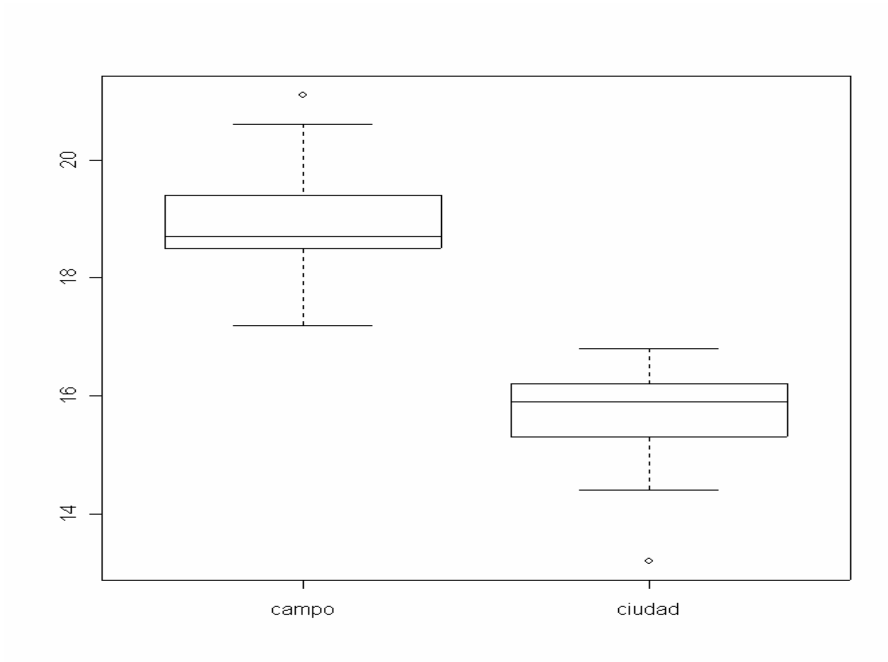
En una prueba de rendimiento y consumo de gasolina se probaron 13 vehículos, durante 300 millas, en condiciones de tránsito en ciudad y en el campo; de lo anterior se obtuvieron los siguientes datos en milla por galón:

Ciudad	16,2	16,7	15,9	14,4	13,2	15,3	16,8	16,0	16,1	15,3	15,2	15,3	16,2
Campo	19,4	20,6	18,3	18,6	19,2	17,4	17,2	18,6	19,0	21,1	19,4	18,5	18,7

Realice un análisis descriptivo de los datos que incluya un gráfico y las medidas descriptivas adecuadas para determinar la relación en el rendimiento y consumo de gasolina para la ciudad y el campo.

En este caso lo más apropiado para realizar el análisis es construir un boxplot para cada conjunto de datos. El mismo nos permite visualizar en un solo dibujo una serie de medidas descriptivas básicas para describir el comportamiento de los datos. Es importante destacar que las unidades de medición de los grupos deben ser las mismas.

Siguiendo los pasos descritos anteriormente para la construcción del boxplot se obtiene:



De aquí se puede concluir lo siguiente:
En el campo el consumo medio de gasolina resultó mayor al de la ciudad, lo que se aprecia en los valores de las medianas (Ciudad:15,9, Campo:18,7).
La variabilidad de ambos grupos es semejante, lo que se observa en el ancho de las cajas, que representa el rango intercuantil.
En cuanto a la simetría se tiene que para el grupo del campo la distribución es asimétrica positiva mientras que para el grupo de la ciudad se observa asimetría negativa.
Por otra parte se observa un dato atípico en el campo y otro en la ciudad.

Probabilidades

La teoría de la probabilidad es un modelamiento matemático del fenómeno del azar o aleatoriedad. Entre los conceptos básicos se tienen:

- **Experimento**

Es toda acción o proceso que produce resultados bien definidos.
Ejemplos:

Experimento	Resultado
Lanzar una moneda	Cara o Sello
Seleccionar una parte para inspeccionarla	Defectuosa o no defectuosa
Lanzar un dado	1, 2, 3, 4, 5, 6
Jugar un partido de fútbol	Ganar, perder, empatar

- **Experimentos aleatorios**

Son aquellos experimentos en los cuales los resultados no son esencialmente los mismos a pesar de que las condiciones sean aproximadamente idénticas. Diremos que un experimento es aleatorio si se verifican las siguientes condiciones:
- Se puede repetir indefinidamente, siempre en las mismas condiciones.
- Antes de realizarlo, no se puede predecir el resultado que se va a obtener.

- **Espacio muestral**

Es aquel conjunto que contiene a todos los resultados de un experimento aleatorio, puede ser finito o infinito y discreto o continuo.
Por ejemplo, si se lanza un dado, el espacio muestral de todos los resultados es $\{1,2,3,4,5,6\}$.

- **Suceso o evento**

Es cualquier subconjunto de resultados contenido en el espacio muestral. Ejemplo: Si se lanza una moneda al aire 2 veces, el hecho de que sólo resulte cara es un suceso del espacio muestral.

- **Tipos de sucesos**

- Suceso cierto o seguro: es aquel que siempre ocurre.
- Suceso imposible: es aquel que no puede ocurrir.
- Sucesos mutuamente excluyentes: son aquellos que no pueden ocurrir simultáneamente, por lo que no tienen elementos comunes.

Ejemplo: lanzar una moneda al aire, el obtener cara o sello es un suceso mutuamente excluyente.
- Sucesos independientes: son aquellos donde la ocurrencia de uno no afecta la ocurrencia del otro.

Nota: dos eventos cuyas probabilidades son diferentes de cero no pueden ser simultáneamente independientes y mutuamente excluyentes entre sí. Si ocurre uno de los eventos mutuamente excluyentes, la probabilidad de que suceda el otro es cero y en consecuencia son dependientes.
- Sucesos complementarios: dos sucesos son complementarios si la no aparición de uno de ellos obliga a que ocurra el otro.

Ejemplo: si A es el suceso de sacar un número par con un dado, el complemento es sacar un número impar.

- Sucesos colectivamente exhaustivos: los eventos A_1, A_2, ..., A_n son colectivamente exhaustivos si la unión de ellos da el espacio muestral.

❖ Probabilidad

Es la medida de la oportunidad con la que podemos esperar que un suceso ocurra, asignando un número entre 0 y 1 a dicha medida.

❖ Tipos de enfoques

- Enfoque clásico o a priori de la probabilidad: este enfoque se aplica cuando se usa la hipótesis de resultados igualmente probables como la base para asignar probabilidades. Si un suceso puede ocurrir en h maneras diferentes de un número total de n maneras posibles (todos igualmente factibles), entonces la probabilidad del suceso es:

$$p = \frac{h}{n}$$

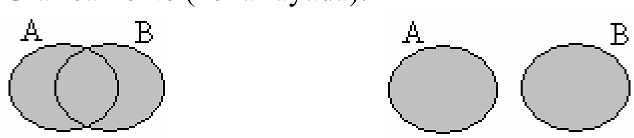
- Enfoque de frecuencia relativa: en este enfoque se utilizan datos pasados obtenidos en observaciones empíricas, teniéndose en cuenta la frecuencia con que ha ocurrido un suceso en el pasado y se estima la probabilidad de que vuelva a ocurrir a partir de estos datos históricos.
Un problema al aplicar este enfoque es el de hacer estimaciones con un número insuficiente de observaciones.

$$p = \frac{\text{número de veces que el suceso ha ocurrido en el pasado}}{\text{número total de observaciones}}$$

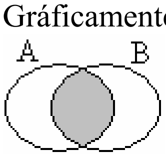
- Enfoque subjetivo: es aquel que se utiliza para asignar una probabilidad a un suceso que no ha ocurrido nunca, según nuestro mejor criterio. Por ejemplo, la probabilidad de que una mujer sea elegida presidente de Venezuela; como no hay datos históricos en que apoyarse, debemos recurrir a nuestra opinión y creencias para hacer una estimación subjetiva.
El tratamiento moderno de la teoría de probabilidad es axiomático en el uso de la teoría de conjuntos.

❖ Algunas relaciones de teoría de conjuntos

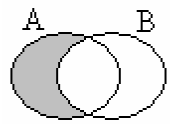
Unión: se llama unión o reunión de dos conjuntos A y B, al conjunto C formado por los elementos que pertenezcan a A o a B.
Notación simbólica: $A \cup B = \{x \mid x \in A \text{ o } x \in B\}$
Gráficamente (zona rayada):



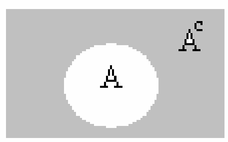
Intersección: se llama intersección de dos conjuntos A y B, al conjunto C formado por los elementos comunes a A y a B.
Notación simbólica: $A \cap B = \{x \mid x \in A \text{ y } x \in B\}$
Gráficamente (zona rayada):



Diferencia: se llama diferencia de dos conjuntos A y B, en este orden, al conjunto C formado por los elementos que pertenecen a A pero no a B.
Notación simbólica: $A - B = \{x \mid x \in A \text{ y } x \notin B\}$
Gráficamente (zona rayada):



Complemento de un conjunto: el complemento de un conjunto A, se denota por A^c o A' y es el conjunto de elementos que pertenecen al conjunto universal pero que no pertenecen a A.
Nota: se supone que todos los conjuntos bajo investigación en cualquier aplicación de una teoría de conjuntos están contenidos en algún conjunto grande fijo denominado conjunto universal o universo de discurso.
Notación simbólica: $A^c = \{x \mid x \in \mathbf{U} \text{ y } x \notin A\}$
Gráficamente (zona rayada):



Ahora bien, un evento no es otra cosa que un conjunto, de modo que se pueden emplear las relaciones y resultados de la teoría básica de conjuntos para estudiar eventos. Los siguientes conceptos de teoría de conjuntos se emplearán para construir nuevos eventos a partir de eventos dados:

a. La unión de dos eventos A y B, denotada por $A \cup B$, es el evento formado por todos los resultados que están ya sea en A o en B o en ambos eventos. (por lo que la unión incluye resultados para los cuales A y B ocurren, así como resultados para los que ocurre solamente un evento).
b. La intersección de dos eventos A y B, denotada por $A \cap B$, es el evento formado por todos los resultados que están en A y B.
c. El complemento de un evento A, denotado por A', es el conjunto de todos los resultados en el espacio muestral que no están contenidos en A.

- **Axiomas y teoremas de probabilidad**

Para cada suceso A.
(a) $0 \leq p(A) \leq 1$
(b) Si A es un suceso seguro: $p(A) = 1$
(c) Si A es un suceso imposible: $p(A) = 0$
(d) Si A' es el complemento de A, entonces $p(A') = 1 - p(A)$
(e) Si $A = A_1 \cup A_2 \cup A_3 \cup ... \cup A_n$ y $A_1, A_2, A_3 ... A_n$ son sucesos mutuamente excluyentes entonces $p(A) = p(A_1) + p(A_2) + p(A_3) + ... + p(A_n)$
En particular, si A es el espacio muestral, $p(A) = p(A_1) + p(A_2) + p(A_3) + ... + p(A_n) = 1$

- **Regla de la suma**

Se usa para determinar la probabilidad de una unión entre eventos.
Si A_1 y A_2 son sucesos cualesquiera entonces:
$p(A_1 \cup A_2) = p(A_1) + p(A_2) - p(A_1 \cap A_2)$.

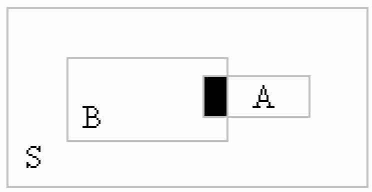
Si A_1 , A_2 y A_3 son sucesos cualesquiera, entonces:
$p(A_1 \cup A_2 \cup A_3) = p(A_1) + p(A_2) + p(A_3) - [p(A_1 \cap A_2) + p(A_2 \cap A_3) + p(A_1 \cap A_3)] + p(A_1 \cap A_2 \cap A_3)$
Para n sucesos tenemos:
$p(A_1 \cup A_2 \cup ... \cup A_n) = \Sigma p(A_k) - \Sigma p(A_j \cap A_k) + \Sigma p(A_i \cap A_j \cap A_k) - ... + (-1)^{n-1} p(A_1 \cap A_2 \cap ... \cap A_n)$
i, j, k = 1, 2, ..., n
(Si los sucesos son mutuamente excluyentes, las probabilidades de las intersecciones son iguales a cero.)

❖ **Probabilidad condicional**

Suponga que B es un evento en un espacio muestral S con p(B) > 0. La probabilidad de que ocurra un evento A una vez que B ha ocurrido, o en otras palabras, la probabilidad condicional de A dado B, escrita p(A | B) se define así:

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)}$$

Como se ilustra en el diagrama de Venn, p(A | B) mide, en cierto sentido, la probabilidad relativa de A con respecto al espacio reducido B



La probabilidad condicional toma en cuenta información sobre la ocurrencia de un evento para predecir la probabilidad de otro.

❖ **Eventos Independientes**

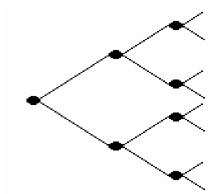
Sean A y B eventos con probabilidades positivas, entonces A y B se dice que son eventos independientes si se cumple que: $p(B \mid A) = p(B)$ o $p(A \mid B) = p(A)$. Extendiendo este concepto, en general se tiene que los eventos A_1, A_2 , …, A_n , se dice que son independientes si y solo si, para cada conjunto de dos o más eventos, la probabilidad de su intersección en el conjunto es igual al producto de las probabilidades de los eventos en el conjunto.

❖ **Regla de la multiplicación**

Se usa para determinar la probabilidad de una intersección entre eventos.
Para 2 sucesos independientes*:*
$p(A_1 \cap A_2) = p(A_1) * p(A_2)$
Para n sucesos independientes*:*
$p(A_1 \cap A_2 \cap ... \cap A_n) = p(A_1) * p(A_2) * ... * p(A_n)$
Para 2 sucesos dependientes*:*
$p(A_1 \cap A_2) = p(A_1) * p(A_2|A_1)$
Para 3 sucesos dependientes*:*
$p(A_1 \cap A_2 \cap A_3) = p(A_1) * p(A_2|A_1) * p(A_3|A_1 \cap A_2)$

❖ **Diagrama de árbol**

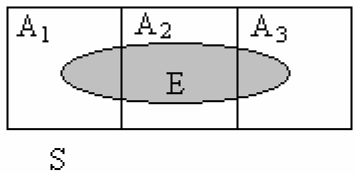
Un diagrama de árbol es una representación gráfica útil para organizar cálculos que abarcan varias etapas. Cada segmento en el árbol es una etapa del problema. Las probabilidades escritas cerca de las ramas son las probabilidades condicionales del experimento. Ejemplo:



La cantidad de ramas van aumentando progresivamente, según la cantidad de etapas.

❖ Particiones, Probabilidad Total y Teorema de Bayes

Supongamos que un conjunto S es la unión de subconjuntos mutuamente excluyentes A_1, A_2, A_3 ... A_n , (son particiones del conjunto S). Además, suponga que E es cualquier subconjunto de S; entonces, como se ilustra en la figura (caso n = 3):



$E = E \cap S = E \cap(A_1 \cup A_2 \cup A_3 \cup ... \cup A_n) = (E \cap A_1) \cup (E \cap A_2) \cup (E \cap A_3) \cup ... \cup (E \cap A_n)$.
Además, los n subconjuntos a la derecha de la ecuación son también mutuamente excluyentes, es decir, forman una partición de E.

Ahora supongamos que S es un espacio muestral y los subconjuntos anteriores A_1, A_2, A_3 ... A_n , E son eventos. Puesto que $E \cap A_k$ son mutuamente excluyentes, se obtiene:

$$p(E) = p(E \cap A_1) + p(E \cap A_2) + p(E \cap A_3) + ... + p(E \cap A_n).$$

Utilizando el teorema de multiplicación para la probabilidad condicional, se obtiene también:

$$p(E \cap A_k) = p(A_k \cap E) = p(A_k) * p(E \mid A_k)$$

Por tanto se llega al siguiente teorema de *probabilidad total*:
Sea E un evento en un espacio muestral S y sean A_1, A_2, A_3 ... A_n eventos mutuamente excluyentes cuya unión es S. Entonces:

$$p(E) = p(A_1) * p(E \mid A_1) + p(A_2) * p(E \mid A_2) + \ldots + p(A_n) * p(E \mid A_n)$$

Por otra parte, se tiene que una fase importante del análisis de probabilidades es su actualización cuando se adquiere información adicional; con frecuencia, comenzamos nuestro análisis con estimados iniciales previos, a priori o anteriores a los eventos específicos de interés. Entonces, con base en fuentes como una muestra, obtenemos cierta información adicional sobre los eventos. Con esa nueva información modificamos los valores de las probabilidades previas mediante el cálculo de probabilidades actualizadas a las que llamamos probabilidades posteriores o a posteriori. A través del teorema de Bayes podemos obtener la probabilidad condicional de un evento cuando mediante el efecto tratamos de determinar la probabilidad de la causa.
Supóngase que A_1, A_2, ..., A_n son sucesos mutuamente excluyentes cuya unión es el espacio muestral S. Además, suponga que E es cualquier evento en S, donde p (E) > 0. La probabilidad condicional de A_k dado que el evento E ha ocurrido está expresada por:

$$p(A_k \mid E) = \frac{p(A_k \cap E)}{p(E)} = \frac{p(A_k) * p(E \mid A_k)}{\sum_{k=1}^{n} p(A_k) * p(E \mid A_k)} = \frac{p(A_k) * p(E \mid A_k)}{p(A_1) * p(E \mid A_1) + p(A_2) * p(E \mid A_2) + \cdots + p(A_n) * p(E \mid A_n)}$$

Distribuciones de Probabilidad

Con anterioridad se ha visto que las distribuciones de frecuencia han resultado ser una forma útil de resumir las variaciones en los datos observados, ahora se trabajará con distribuciones de probabilidad. La diferencia entre estas dos distribuciones consiste en que una distribución de frecuencias es un listado de las frecuencias observadas de todos los resultados de un experimento que se presentaron realmente cuando se efectuó el experimento, mientras que una distribución de probabilidad es un listado de las probabilidades de todos los posibles resultados que *podrían* obtenerse si el experimento se llevara a cabo. También, las distribuciones de probabilidad pueden basarse en consideraciones teóricas, en una estimación subjetiva de la posibilidad de ciertos resultados o en la experiencia.

- **Tipos de distribuciones de probabilidad**

1) Distribuciones de probabilidad discretas: en este tipo de distribución está permitido tomar sólo un número limitado de valores. Por ejemplo, la probabilidad de que usted haya nacido en un mes dado es discreta puesto que sólo hay 12 posibles valores (los meses del año).

2) Distribuciones de probabilidad continuas: en este tipo de distribución la variable que se está considerando puede tomar cualquier valor dentro de un intervalo dado.

- **Variables Aleatorias**

Una primera interpretación es que una variable es aleatoria si se toma diferentes valores como resultado de un experimento aleatorio. Al precisar más el concepto podemos establecer que una variable aleatoria es una función de valor real que tiene como dominio el espacio muestral asociado a un experimento aleatorio, esto quiere decir que es una función a través de la cual a cada punto de un espacio muestral le asignamos un número. Comúnmente se denota por una letra mayúscula **X**; **Y**; **Z**.

Variable aleatoria es toda función $\mathbf{X}: E \rightarrow \mathbf{IR}$ $e \rightarrow \mathbf{X}(e) = x_e$ que atribuye un único número real x_e , a cada suceso elemental e

Ejemplo:
Supóngase que se lanza una moneda 2 veces de tal forma que el espacio muestral es: δ = {CC, CS, SC, SS}, represéntese por **X** el número de caras que pueden resultar.
Con cada punto muestral podemos asociar un número para **X**, así, en el caso de CC (dos caras), **X** = 2, en tanto que para SC (una cara), **X** = 1, etc.

Tipos de variables aleatorias

1) Variable aleatoria discreta: si la variable toma un número contable de valores en un intervalo. Por ejemplo, el número de puntos que muestra la cara superior de un dado después de su lanzamiento; el número de clientes que llegan a una hora a un banco en solicitud de servicio, etc.

$$\mathbf{X} : E \rightarrow \mathrm{IN}$$

2) Variable aleatoria continua: si la variable toma cualquier valor dentro de un intervalo dado. Ejemplo: el peso de los seres humanos.

$$\mathbf{X} : E \rightarrow IR$$

Distribución de probabilidad de variables aleatorias discretas

El término distribución de probabilidad, se refiere a la colección de valores de la variable aleatoria junto a la distribución de probabilidad entre éstos; implica la existencia de la función de probabilidad y de la función de distribución acumulativa de **X**.

- **Función de probabilidad o función de masa f(x)**

La variable aleatoria **X** representa los resultados de un espacio muestral en forma tal que por P(**X** = x) se entenderá la probabilidad de que **X** tome el valor de "x", de esta forma, al considerar los valores de una variable aleatoria es posible desarrollar una función que determina la probabilidad de que "x" asuma un valor determinado de una variable aleatoria discreta. Esta función recibe el nombre de función de probabilidad, algunos autores la llaman también distribución de probabilidad.

Definición:

Sea **X** una variable aleatoria discreta, se llamará P(X = x) = f(x) función de probabilidad de la variable aleatoria, si satisfacen las siguientes propiedades:

1. $f(x) \geq 0$ o $p(x) \geq 0$ $\forall x \in \mathbf{X}$
2. $\sum_x f(x) = 1$ $\sum_x p(x) = 1$ donde la suma se toma sobre los valores posibles de x

- **Función de distribución acumulativa F(x)**

La función de distribución acumulativa de la variable aleatoria **X** o como algunos autores indican, función de distribución, es la probabilidad de que **X** sea menor o igual a un valor específico de "x" y está dada por:

$F(x) = p(\mathbf{X} \leq x) = \sum_{x_i \leq x} p(x_i)$

Esta función representa la suma de las probabilidades puntuales hasta el valor "x" de **X** inclusive, en general, la función de distribución acumulativa F(x) de una variable aleatoria discreta es una función no decreciente de los valores de **X**, de tal manera que:

1. $0 \leq F(x) \leq 1$ para cualquier x
2. $F(x_i) \geq F(x_j)$ si $x_i \geq x_j$
3. $p(\mathbf{X} > x) = 1 - F(x)$

Si **X** únicamente toma un número infinito de valores $x_1, x_2, ..., x_n$ entonces la función de distribución está dada por:

$$F(x) = \begin{cases} 0 & -\infty < x < x_1 \\ f(x_1) & x_1 \leq x < x_2 \\ f(x_1) + f(x_2) & x_2 \leq x < x_3 \\ f(x_1) + f(x_2) + f(x_3) & x_3 \leq x < x_4 \\ \vdots & \vdots \\ f(x_1) + \cdots + f(x_n) & x_n \leq x < \infty \end{cases}$$

- **Esperanza o valor esperado de una variable aleatoria discreta**

A la esperanza de **X** frecuentemente se le llama media de **X** y se denota por μ_x o simplemente μ cuando se sobreentiende la variable aleatoria determinada. La media o esperanza de **X** da un valor típico o promedio de los valores de **X** y por esta razón se le llama medida de centralización.
Para una variable aleatoria discreta **X** que puede tener los valores $x_1, x_2, ..., x_n$, la esperanza viene dada por la expresión:
$E(\mathbf{X}) = \Sigma\, x * f(x) = x_1 * f(x_1) + x_2 * f(x_2) + x_3 * f(x_3) + ... + x_n * f(x_n)$

Algunos teoremas sobre esperanza

1. Si c es cualquier constante, entonces $E(c\mathbf{X}) = c * E(\mathbf{X})$
2. Si **X**, **Y** son variables aleatorias, entonces $E(\mathbf{X} + \mathbf{Y}) = E(\mathbf{X}) + E(\mathbf{Y})$
3. Si **X**, **Y** son variables aleatorias independientes, entonces $E(\mathbf{XY}) = E(\mathbf{X}) * E(\mathbf{Y})$

- **Varianza y desviación típica de variables aleatorias discretas**

Son medidas de dispersión de los valores de la variable aleatoria alrededor de μ. Si los valores tienden a concentrarse alrededor de la media, la varianza es pequeña, en caso contrario es grande.
$Var(\mathbf{X}) = E[(\mathbf{X} - \mu)^2] = \Sigma\, (x - \mu)^2 * f(x)$
La desviación típica será: $\sqrt{Var(\mathbf{X})}$

Algunos teoremas sobre varianza

1. $Var(\mathbf{X}) = E(\mathbf{X}^2) - [E(\mathbf{X})]^2$
2. Si c es cualquier constante, entonces $Var(c\mathbf{X}) = c^2 * Var(\mathbf{X})$
3. Si **X**, **Y** son variables aleatorias independientes, entonces: $Var(\mathbf{X} + \mathbf{Y}) = Var(\mathbf{X}) + Var(\mathbf{Y})$

Modelos de Distribución de Probabilidad para Variables Aleatorias Discretas

La distribución de probabilidad de una variable discreta es una tabla, gráfica, fórmula o cualquier otro medio que se usa para especificar todos los valores posibles de la variable, junto con sus respectivas probabilidades. Para las variables que se emplean en el estudio de problemas de interés, su distribución se da generalmente mediante una fórmula y para los efectos, trabajaremos con la distribución binomial, la distribución binomial negativa, la distribución geométrica, la distribución hipergeométrica y la distribución de Poisson.

❖ Distribución Binomial

La distribución binomial está ligada a un tipo de experimento llamado ensayo de Bernoulli, en honor a su descubridor Jacob Bernoulli. Este experimento aleatorio que sólo puede concluir de dos maneras distintas mutuamente excluyentes.

Ejemplos de ensayos de Bernoulli son:

1. Seleccionar un artículo para clasificarlo como bueno o defectuoso.
2. Seleccionar una persona para clasificarla como apta o no apta de acuerdo con algún criterio.

Un proceso de Bernoulli, es una secuencia de ensayos de Bernoulli con las características siguientes:

1. Hay una secuencia de n intentos, es decir, el experimento se puede repetir n veces.
2. Los intentos son idénticos y cada uno de ellos puede resultar en uno de dos posibles resultados: éxito o fracaso. El éxito tiene una probabilidad "p" y el fracaso una probabilidad "q" de ocurrir ($q = 1 - p$). Para nuestros propósitos, se tomará como éxito al aspecto en el cual centramos nuestra atención.
3. La probabilidad de éxito y de fracaso permanece constante durante el proceso.
4. Los ensayos son independientes, es decir, el resultado de cualquier ensayo particular no es afectado por el resultado de cualquier otro ensayo.

Los experimentos de este tipo siguen la que se ha denominado distribución binomial, por lo que la variable aleatoria **X** con distribución binomial debe ser de la forma:

X = número de éxitos en los n ensayos de Bernoulli.

- **Función de probabilidad**

La probabilidad de que el suceso ocurra "x" veces en n intentos cumpliendo con las condiciones anteriores, está dada por la función de probabilidad:

$$f(x) = p(\mathbf{X} = x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & \text{en donde } x = 0, 1, 2, ..., n \\ 0 & \text{para cualquier otro valor} \end{cases}$$

donde:

- f(x) se denomina distribución binomial o distribución de Bernoulli
- p es la probabilidad de éxito.
- q es la probabilidad de fracaso.
- n es el número de ensayos.
- $\binom{n}{x} = C_{n,x} = \dfrac{n!}{x! * (n - x)!}$

- **Parámetros de la distribución binomial**

Los parámetros de la distribución son "n" y "p", la notación que también se puede seguir es: **X** ~> B(n, p) o también B(x; n, p).

- **Media y varianza para una distribución binomial**

E(**X**) = μ_x = np
Var(**X**) = npq

❖ **Distribución Binomial Negativa**

Sea un escenario binomial en que se observa una secuencia de ensayos independientes; la probabilidad de éxito en cada ensayo es constante e igual a p. Ahora bien, si en lugar de fijar el número de ensayos en n y observar el número de éxitos, supóngase que se continúan los ensayos hasta que han ocurrido exactamente k éxitos. En este caso, la variable aleatoria es el número de ensayos necesarios para observar k éxitos, esta situación lleva a lo que se conoce como la distribución binomial negativa.
La variable aleatoria **X** con distribución binomial negativa debe ser de la forma:

X = número de ensayos necesarios para observar k éxitos

Las condiciones que deben presentarse son:

- El experimento consta de una secuencia de intentos independientes.
- Cada intento puede resultar en un éxito o un fracaso.
- La probabilidad de éxito es constante de un intento a otro.
- El experimento continúa (los intentos se ejecutan) hasta que un total de k éxitos se haya observado

- **Función de probabilidad**

La función de probabilidad de la variable aleatoria **X** está dada por

$$f(x) = \begin{cases} \binom{x-1}{k-1} p^k (1-p)^{x-k} & \begin{array}{l} x = k, k+1, k+2, \ldots \\ k = 1, 2, \ldots \\ 0 \le p \le 1 \end{array} \\ 0 & \text{para cualquier otro valor} \end{cases}$$

- **Parámetros de la distribución Binomial Negativa**

Los parámetros de la distribución binomial negativa son *k* y *p*, en donde *k* no necesita ser un entero. Si es así, la distribución se conoce como distribución de Pascal, misma que se interpreta como el tiempo que hay que esperar para que ocurra el *k* éxito. Si *k* no es entero, la función de probabilidad se escribe de manera tal que se involucre a la función gama, situación que va más allá de los objetivos del curso.

- **Media y Varianza para una distribución Binomial Negativa**

E(**X**) = $\frac{k}{p}$ Var(**X**) = $\frac{k(1-p)}{p^2}$

❖ Distribución Geométrica

Considere ensayos repetidos de un experimento de Bernoulli con probabilidad p de éxito y q = 1 - p de fracaso. Sea **X** el número de veces que el experimento debe ser repetido hasta obtener finalmente un éxito. Una distribución de esta naturaleza se conoce como distribución geométrica y es un caso particular de la distribución binomial negativa, en la cual k = 1. Un ejemplo de esta distribución sería disparar continuamente a un objetivo hasta alcanzarlo finalmente.

La variable aleatoria **X** con distribución geométrica debe ser de la forma:

X = número de intentos hasta que ocurra el primer éxito

- **Función de probabilidad**

La función de probabilidad de la variable aleatoria **X** está dada por

$$f(x) = \begin{cases} p(1-p)^{x-1} & x = 1, 2, ... \\ & 0 \leq p \leq 1 \\ 0 & \text{para cualquier otro valor} \end{cases}$$

- **Parámetro de la distribución Geométrica**

El parámetro de la distribución geométrica es p

- **Media y Varianza para una distribución Geométrica**

$$E(\mathbf{X}) = \frac{1}{p} \qquad \mathrm{Var}(\mathbf{X}) = \frac{q}{p^2}$$

❖ Distribución Hipergeométrica

La distribución binomial se basa en el supuesto de que la población es infinita y de que la probabilidad de éxito permanece constante, lo cual se consigue en tales poblaciones o cuando se toman muestras con reemplazo en poblaciones finitas. Cuando la población es finita y el muestreo se hace sin reemplazo, la probabilidad cambiará para cada nueva observación. En tales circunstancias, se tendrá una distribución de probabilidad que se llama distribución hipergeométrica. La muestra debe estar formada por 2 grupos de individuos u objetos (aquellos que poseen la característica objeto de estudio y los que no la poseen).

La variable aleatoria **X** con distribución hipergeométrica debe ser la forma:

X = número de éxitos en los n ensayos (muestreo sin repetición).

- **Función de probabilidad**

Los valores de probabilidad asociados a esta variable con distribución hipergeométrica están dados por la función de probabilidad

$$f(x) = p(\mathbf{X} = x) = \begin{cases} \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} & \text{en donde } x = 0, 1, 2, ..., n \quad x \le M \quad n - x \le N - M \\ 0 & \text{para cualquier otro valor} \end{cases}$$

donde:

- N es el número de individuos donde debe hacerse el muestreo.
- M es el éxito en la población.
- n es el tamaño de la muestra.
- x es el número de éxitos en la muestra.

- **Parámetros de la distribución hipergeométrica**

Los parámetros de la distribución son “n”, “N”, “M”, la notación que también se puede seguir es H(x; N, M, n).

- **Media y varianza para una distribución hipergeométrica**

$E(\mathbf{X}) = n * \frac{M}{N}$

$Var(\mathbf{X}) = \frac{N-n}{N-1} * n * \frac{M}{N} * \left(1 - \frac{M}{N}\right)$

- **Uso de la distribución binomial en la aproximación de la distribución hipergeométrica**

Si se cumple que $n \le 0,05 * N$, podemos utilizar la distribución binomial como una aproximación a la hipergeométrica si se hace un muestreo sin reposición de una población finita.

❖ Distribución de Poisson

Otra familia de distribuciones de probabilidad, también de naturaleza discreta, es la llamada distribución de Poisson, llamada así por Simeón Poisson quien la descubrió en 1837. Esta distribución ha resultado aplicable a muchos procesos en los que ocurren determinados sucesos por unidad de tiempo, espacio, volumen, área, etc. Son casos de este tipo los siguientes:

- Número de accidentes por semana en una autopista.
- Número de personas que llegan en una hora a un banco en solicitud de servicios.
- Número de errores por página que comete una mecanógrafa.
- Número de quejas por pasajero en una línea aérea.
- Número de bacterias por cm^2 en un cultivo.

La variable aleatoria **X** con distribución de Poisson debe ser de la forma:

X = número de éxitos de que ocurra el suceso por unidad de tiempo, espacio, etc.

Las condiciones que deben presentarse son:

- La probabilidad de que ocurra el suceso es constante para dos intervalos de tiempo o espacio cualquiera.
- La aparición de un suceso en cualquier intervalo es independiente de su aparición en cualquier otro intervalo.

- **Función de probabilidad**

La función de probabilidad de la variable aleatoria **X** está dada por

$$f(x) = p(\mathbf{X} = x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} & \text{en donde } x = 0, 1, 2, ..., n \quad \lambda > 0 \\ 0 & \text{para cualquier otro valor} \end{cases}$$

donde:

- λ es el promedio de ocurrencia del suceso en la unidad de tiempo, espacio, volumen, etc.
- e es el número de Euler ($e \approx 2{,}71828$)
- x es el número de veces que ocurre el suceso.

- **Parámetro de la distribución de Poisson**

El parámetro de la distribución es λ, la notación que también se puede seguir es $p(x; \lambda)$

- **Media y Varianza para una distribución de Poisson**

$E(\mathbf{X}) = \lambda$
$Var(\mathbf{X}) = \lambda$

- **Uso de la distribución de Poisson para aproximar a la distribución binomial**

En una distribución binomial, si "n" es grande y "p" está cerca de cero, el suceso se llama suceso raro. En la práctica algunos estadísticos consideran que un suceso es raro si el número de pruebas es al menos 100 ($n \geq 100$) mientras que $np < 20$. En tales casos la distribución binomial se aproxima mucho a la distribución de Poisson con $\lambda = np$.
Algunos autores toman como condición:

1. $n \geq 20$ y $p \leq 0{,}05$
2. $n \geq 50$ y $np \leq 5$
3. $n > 30$ y $p \leq 0{,}1$
4. $n \geq 20$ y $p \leq 0{,}1$

Ejercicios

1) Establecer cuáles de estos datos son discretos y cuáles continuos:
   a. Temperaturas medidas en un laboratorio cada media hora.
   b. Ingresos anuales de los profesores de educación media.
   c. Longitudes de 100 tornillos producidos en una empresa.
   d. Número de estudiantes en un aula de un liceo.

Soluciones: continuo, continuo, continuo, discreto.

2) Clasificar cada una de las siguientes variables:
   a. Distancia diaria recorrida por cada estudiante para ir de su casa a la universidad.
   b. Tiempo que requiere un estudiante para responder a un examen.
   c. Llamadas que llegan a la central telefónica de la USB en un día.
   d. Preferencia por cierta marca de refresco.
   e. Color del cabello de las estudiantes que toman el curso de estadística en el trimestre.
   f. Número de acciones vendidas en un día en la Bolsa de Valores.
   g. Vida media de los tubos de televisión producidos por una fábrica

Soluciones: cuantitativa-continua, cuantitativa-continua, cuantitativa-discreta, cualitativa-discreta, cualitativa-discreta, cuantitativa-discreta, cuantitativa-continua.

3) Se ha hecho un estudio para determinar la preferencia de una marca especial de detergente por parte de las amas de casa. Entre las 50 amas de casa entrevistadas, 30 dijeron que preferían esta marca.
   a. ¿Qué constituye la muestra?.
   b. ¿Qué constituye la población?.
   c. ¿Cuál es la proporción, dentro de la muestra, de las amas de casa que prefieren la marca del detergente?

Soluciones: el conjunto de respuestas que dieron las 50 amas de casa., el conjunto formado por las posibles respuestas de las amas de casa., 0,6.

4) En una fiesta, el 50% de los invitados son hombres. De todos los hombres de la fiesta, el 40% son calvos y de ellos el 50% habla inglés. Si 4 calvos hablan inglés. ¿Cuántas mujeres hay en la fiesta?.
Solución: 20 mujeres.

5) Efectuar dos descuentos consecutivos, primero de un 10% y luego de un 20%, es equivalente a efectuar un solo descuento de...
Solución: 28%

6) Si Pedro tuviera un 15% menos de la edad que tiene, tendría 34 años. Hallar su edad actual.
Solución: 40 años.

Ejercicios (Estadística Descriptiva)

1) La tabla muestra una distribución de frecuencias de la duración de 400 tubos de radio comprobados en la L & M Tube Company.

Duración (horas)	Número de tubos
[300-400)	14
[400-500)	46
[500-600)	58
[600-700)	76
[700-800)	68
[800-900)	62
[900-1000)	48
[1000-1100)	22
[1100-1200)	6

n = 400

Completar la tabla para luego determinar:

a. Límite superior de la quinta clase.
b. Límite inferior de la octava clase.
c. Marca de clase de la séptima clase.
d. Tamaño del intervalo de clase.
e. Frecuencia de la cuarta clase.
f. Frecuencia relativa de la sexta clase.
g. Porcentaje de tubos cuya duración es menor a las 600 horas.
h. Porcentaje de tubos cuya duración es mayor o igual a 900 horas.
i. Porcentaje de tubos cuya duración es al menos de 500 horas pero menor de 1000 horas.
j. Construir un histograma y un polígono de frecuencias.
k. Construir un histograma y un polígono de frecuencias relativas.
l. Construir una ojiva porcentual.
m. Estimar el porcentaje de tubos con duraciones de menos de 560 horas.
n. Estimar el porcentaje de tubos con duraciones de 970 o más horas.
o. Estimar el porcentaje de tubos con duraciones entre 620 y 890 horas.

Sol:

a. 800 horas
b. 1000 horas
c. 950 horas
d. 100 horas
e. 76 tubos
f. 62/400 = 0,1555 ó 15,55%
g. 29,5 %
h. 19,00%
i. 78,00%
m. 23,70%
n. 10,60%
o. 46,15%

2) Los diámetros interiores de las arandelas producidas por una compañía pueden medirse con una aproximación de milésimas de pulgada. Si las marcas de clase de distribución de frecuencias

de estos diámetros vienen dadas en pulgadas por los números: 0,321; 0,324; 0,327; 0,.330; 0,333 y 0,336.

Hallar:	Sol:
a. El tamaño de intervalo de clase.	a. 0,003 pulgadas
b. Los límites de clase.	b. 0,3195; 0,3225; 0,3255; ... ; 0,3375 pulgadas

3) La siguiente tabla muestra los diámetros en pulgadas de nuestra muestra de 60 cojines de bolas fabricados por una compañía.

0,738	0,729	0,743	0,740	0,736	0,741	0,735	0,731	0,726	0,737
0,728	0,737	0,736	0,735	0,724	0,733	0,742	0,736	0,739	0,735
0,745	0,736	0,742	0,740	0,728	0,738	0,725	0,733	0,734	0,732
0,733	0,730	0,732	0,730	0,739	0,734	0,738	0,739	0,727	0,735
0,735	0,732	0,735	0,727	0,734	0,732	0,736	0,741	0,736	0,744
0,732	0,737	0,731	0,746	0,735	0,735	0,729	0,734	0,730	0,740

Construir una tabla de distribución de frecuencias de los diámetros y grafique:

a. Un histograma.
b. Un polígono de frecuencias relativas.
c. Una ojiva y una ojiva porcentual.

Determinar:

a. El porcentaje de cojinetes de bolas que tienen diámetros superiores a 0,732 pulgadas
b. El porcentaje de cojinetes de bolas que tienen diámetros no superiores a 0,736 pulgadas
c. El porcentaje de cojinetes de bolas que tienen diámetros entre 0,730 y 0,738 pulgadas.

4) A continuación, se ofrece una distribución de frecuencia del peso de 150 personas que utilizaron un elevador cierto día.

Clase	f_i
[75-90)	10
[90-105)	11
[105-120)	23
[120-135)	26
[135-150)	31
[150-165)	23
[165-180)	9
[180-195)	9
[195-210)	6
[210-225)	2

Construya un histograma con esos datos.

5) Homero Willis, capitán de un barco pesquero de Salter Path (North Carolina) tiene la creencia de que la pesca mínima para recuperar la inversión debe ser de 5000 libras por viaje. A continuación tenemos los datos de una muestra de la pesca de 20 salidas al mar que el barco de Homero ha hecho recientemente:

6500	6700	3400	3600	2000
7000	5600	4500	8000	5000
4600	8100	6500	9000	4200
4800	7000	7500	6000	5400

Tomando 4 clases y con una amplitud de 2000, construya una tabla de distribución de frecuencias y una ojiva que le ayude a contestar las siguientes preguntas:

a. ¿Aproximadamente qué proporción de los viajes recupera y sobrepasa la inversión según Homero?
b. ¿Qué pescas del barco de Willis superan el 20%?

Sol:

a. Aproximadamente 0,675
b. Aproximadamente 4300 libras.

6) Antes de construir una presa sobre un río, se efectuaron una serie de pruebas para medir el flujo de agua que pasa por el lugar de la presa. Los resultados de las pruebas se usaron para preparar la siguiente distribución de frecuencia:

Flujo del río (miles de galones por minuto)	frecuencia
[1001-1051)	7
[1051-1101)	21
[1101-1151)	32
[1151-1201)	49
[1201-1251)	58
[1251-1301)	41
[1301-1351)	27
[1351-1401)	11

n = 246

a. Con los datos de la tabla anterior construya una distribución de frecuencias
b. Construya una ojiva relativa
c. Por medio de la ojiva relativa, estime qué proporción del flujo ocurre en menos de 1300 galones por minuto.
   Sol: 0,842

7) Nora Velarde, asesora de una pequeña empresa de corretaje, intenta diseñar programas de inversión que fuesen atractivos para jubilados. Ella sabe que si un inversionista potencial pudiera obtener un cierto nivel de intereses, estaría dispuesto a invertir su capital, pero debajo de un cierto nivel de intereses, no estaría dispuesto a hacerlo. De un grupo de 50 sujetos, Nora obtuvo los datos siguientes con respecto a los diferentes niveles de réditos requeridos por cada individuo para que pueda invertir 1000 dólares:

Punto de diferencia ($)	f_i
[70 – 75)	2
[75 – 80)	5
[80 – 85)	10
[85 – 90)	14
[90 – 95)	11
[95 – 100)	3
[100 – 105)	3
[105 – 110)	2

n = 50

a. Construya la distribución de frecuencia acumulativa.
b. Grafique la distribución de la parte (a) convirtiéndola en ojiva de frecuencia relativa.

8) En la oficina de un diario, el tiempo que se tardan en imprimir la primera plana fue registrado durante 50 días. A continuación se transcriben los datos, aproximados a décimas de minuto:

20,8	22,8	21,9	22,0	20,7	20,9	25,0	22,2	22,8	20,1
25,3	20,7	22,5	21,2	23,8	23.3	20,9	22,9	23,5	19,5
23,7	20,3	23,6	19,0	25,1	25,0	19,5	24,1	24,2	21,8
21,3	21,5	23,1	19,9	24,2	24,1	19,8	23,9	22,8	23,9
19,7	24,2	23,8	20,7	23,8	24,3	21,1	20,9	21,6	22,7

a. Construya con los datos una tabla de distribución de frecuencia, usando intervalos de 0,8 minutos.
b. Construya un polígono de frecuencias.
c. Construya una ojiva.
d. Por medio de la ojiva estime que porcentaje de las veces la primera plana del periódico puede imprimirse en menos de 24 minutos.
   Sol: aproximadamente un 75,5%.

9) Un agente de seguros tiene datos sobre la cantidad mensual de pólizas que vendió en los 3 últimos años. Sus datos los ha arreglado en la siguiente distribución de frecuencia:

Ventas mensuales	f_i
[1000-1150)	1
[1150-1300)	3
[1300-1450)	6
[1450-1600)	4
[1600-1750)	8
[1750-1900)	9
[1900-2050)	3
[2050-2200)	2

a. Construya una distribución de frecuencia relativa.
b. Construya un histograma de frecuencia relativa y un polígono de frecuencias relativas.
c. Construya una ojiva porcentual.

10) Una compañía fabrica 15 productos básicos. La compañía conserva el registro del número de cada producto fabricado por mes, a fin de examinar los niveles relativos de producción. Los registros muestran que los siguientes números de cada producto fueron fabricados por ella en el último mes de 20 días laborables:

9897	10052	10028	9722	9908
10098	10587	9872	9956	9928
10123	10507	9910	9992	10237

a. Construya una tabla de distribución de frecuencias tomando en cuenta: 5 intervalos, $a = 200$, $l_i = 9700$.
b. Construya una ojiva para responder.
   (1) ¿En cuántos de sus artículos la producción rebasó el punto de equilibrio de 10.000 unidades?
   (2) ¿Qué nivel de producción alcanzó el 25% de sus productos en el mes?

Sol:
(1) Entre 7 y 8 productos
(2) Aproximadamente 9919 unidades.

11) En una población estudiada, hay 2000 mujeres y 8000 hombres. Si queremos seleccionar una muestra de 250 individuos en dicha población. ¿Cuántos deberán ser mujeres para que la muestra sea considerada representativa?
Sol: 50.

12) Si los siguientes grupos de edad están incluidos en las proporciones indicadas. ¿Cuántos individuos de cada grupo deben ser incluidos en una muestra de 2500 personas para que la muestra sea representativa?

Grupo de edad	h_i
[12-18)	0,1300
[18-24)	0,3400
[24-30)	0,2400
[30-36)	0,1800
Más de 36	0,1100

13) La Kawahondi Computer Company recopiló datos referentes al número de entrevistas que necesitaban sus 40 vendedores para realizar una venta. A continuación se dan una distribución de frecuencias absolutas y relativas del número de entrevistas que se necesitan por vendedor para lograra una venta. Anote los datos faltantes:

Números de entrevistas	f_i	h_i
[1-11)	?	0,0500
[11-21)	0	?
[21-31)	2	?
[31-41)	?	?
[41-51)	?	0,1500
[51-61)	?	0,2000
[61-71)	5	?
[71-81)	?	0,0000
[81-91)	5	?
[91-101)	?	0,0000

14) Bill Bissey, vicepresidente del Bank One de Indianápolis, lleva un control de la aprobación de préstamos para el desarrollo de empresas locales. A lo largo de los cinco últimos años el préstamo de mayor cuantía fue de 1,2 millones de dólares y el más pequeño de 10000 dólares. Desea construir una tabla de frecuencias con 10 clases.
¿Cuáles serían los límites de las clases?
¿Cuál sería el intervalo de clase?

15) Mr. Bissey, el vicepresidente del Bank One de Indianápolis, lleva también un registro de las cuentas de ahorro personal. Los saldos de las 40 nuevas cuentas que se abrieron el último mes fueron:

179,80	890,00	712,10	415,00
112,17	1200,00	293,00	602,02
1150,00	1482,00	579,00	312,52
100,00	695,15	287,00	1175,00
1009,10	952,51	1112,52	783,00
1212,43	510,52	1394,05	1390,00
470,53	783,00	1101,00	666,66
780,00	793,10	501,01	1555,10
352,00	937,01	711,11	1422,03
1595,10	217,00	1202,00	1273,01

Construir una tabla de distribución de frecuencias con siete clases.

16) Suponga que usted es el estadístico oficial de líneas aéreas KLM y que el presidente del consejo de administración le ha pedido que recoja y organice datos relativos a las operaciones de vuelo. Su interés principal a partir de los valores diarios se centra en la variable de número de pasajeros.

Ha obtenido estos datos de los diarios de vuelo de los últimos 50 días y ha reflejado esta información:

68	72	50	70	65	83	77	78	80	93
71	74	60	84	72	84	73	81	84	92
77	57	70	59	85	74	78	79	91	102
83	67	66	75	79	82	93	90	101	80
79	69	76	94	71	97	95	83	86	69

a. Construir la tabla de distribución de frecuencias.
b. Construir un histograma y un polígono de frecuencias.
c. Construir una ojiva.

17) A continuación se indican las pérdidas y ganancias, en millones de dólares, de las 50 mayores empresas (por ventas) de la lista de 500 de Fortune en 1992. El valor más bajo es una pérdida de 4453 millones de dólares y el más alto una ganancia de 5600 millones. Construir una tabla de frecuencias con a = 1500, $l_i = -4500$.

-4453.00	-1484.00	-732.00	2056.00	2636.00	454.00	184.00	-387.00	17.00
5600.00	20.00	-617.00	97.00	120.00	-404.00	535.00	423.00	63.00
-2258.00	-1021.00	1154.00	939.00	311.00	3006.00	-142.00	1461.00	
-2827.00	1080.00	-1086.00	460.00	258.00	842.00	1403.00	308.00	
1294.00	454.00	601.00	73.00	1293.00	709.00	-273.00	97.00	
-795.00	-578.00	1681.00	505.00	1567.00	1773.00	368.00	755.00	

18) Usted, en su calidad de consultor económico privado, considera necesario leer detenidamente The Wall Street Journal para estar al corriente en su campo profesional. En un reciente informe del WSJ se facilitaban los siguientes datos como porcentajes de ejecutivos en 42 de las mayores empresas de Estados Unidos que tenían problemas de abuso de medicamentos:

5,9	8,8	14,3	8,3	9,1	5,1	15,3
17,5	17,3	15,0	9,3	9,9	7,0	16,7
10,3	11,5	17,0	8,5	7,2	13,7	16,3
12,7	8,7	6,5	6,8	13,4	5,5	15,2
8,4	9,8	7,3	10,0	11,0	13,2	16,3

9,1	12,3	8,5	16,0	10,2	11,7	14,2

a. Construir el histograma correspondiente.
b. Construir la distribución de frecuencias
c. Construir el polígono de frecuencias.
d. Construir una distribución de frecuencias acumuladas y su ojiva correspondiente.

19) Según Nielsen Media Research, los cinco programas de TV más vistos a las 8 : 00 PM del 14 de Diciembre de 1997 fueron Congo, The X-Files, Holiday in Your Herat, Ellen Foster y por último Unhappily Ever After. La lista siguiente es una encuesta entre 50 espectadores:

Unhappily	Ellen	Congo	X-Files	X-Files
Ellen	Ellen	X-Files	Ellen	X-Files
Congo	Holiday	Congo	Ellen	X-Files
Ellen	Ellen	X-Files	X-Files	Holiday
Ellen	Ellen	Holiday	Holiday	X-Files
Holiday	X-Files	X-Files	Ellen	Ellen
Holiday	Ellen	Holiday	X-Files	Holiday
Congo	Holiday	Congo	X-Files	Ellen
Congo	Congo	Ellen	X-Files	Holiday
Ellen	Unhappily	Holiday	Congo	Ellen

a. ¿Los datos son cualitativos o cuantitativos?.
b. Determine la tabla de distribución de frecuencias.
c. Trace una gráfica de barras y un diagrama de pastel para estos datos.
d. De acuerdo con la muestra:
   ¿Qué programa tiene la mayor parte del mercado?.
   ¿Cuál lo sigue?.
e. ¿Qué porcentaje tiene el programa Congo?.

20) En Beverage Digest se informa que, con base en las ventas de 1998, las 5 marcas de refrescos que más se vendieron fueron Coke Classic, Diet Coke, Dr.Pepper, Pepsi Cola y Sprite. La lista siguiente proviene de una muestra de 50 compras de esas marcas fue:

Coke Classic	Dr.Pepper	Sprite	Coke Classic	Pepsi Cola
Diet Coke	Diet Coke	Coke Classic	Diet Coke	Coke Classic
Pepsi Cola	Pepsi Cola	Diet Coke	Coke Classic	Coke Classic
Diet Coke	Pepsi Cola	Coke Classic	Diet Coke	Coke Classic
Coke Classic	Coke Classic	Coke Classic	Coke Classic	Pepsi Cola
Coke Classic	Dr.Pepper	Sprite	Sprite	Coke Classic
Sprite	Pepsi Cola	Pepsi Cola	Pepsi Cola	Coke Classic
Dr.Pepper	Coke Classic	Dr.Pepper	Pepsi Cola	Dr.Pepper
Pepsi Cola	Coke Classic	Coke Classic	Pepsi Cola	Pepsi Cola
Diet Coke	Coke Classic	Diet Coke	Pepsi Cola	Sprite

a) Construir la tabla de distribución de frecuencias.
b) Construir una gráfica de barras y un diagrama de pastel.
c) ¿Qué porcentaje de las ventas tienen Pepsi Cola y Coke Classic?.
   Sol: 26% y 38%

21) El Union Bank de Suiza realizó una encuesta internacional para obtener datos acerca de los sueldos por hora de los trabajadores y empleados en todo el mundo. Los trabajadores en los

Ángeles ocuparon el séptimo lugar en el mundo, en términos de mayores salarios por hora. Suponga que los siguientes 25 valores son de sueldos por hora de trabajadores en los Ángeles:

11,50	9,20	15,35	8,00	9,80
11,90	11,75	12,05	14,70	7,05
13,10	6,85	10,25	5,85	13,10
8,40	9,15	11,10	13,65	9,05
9,90	10,05	8,45	13,15	6,65

a. Elabore una tabla de distribución de frecuencias que use clases de 4,00 a 5,99 ; 6,00 a 7,99 y así sucesivamente.
b. Trace un histograma y un polígono de frecuencias.

22) Suponga que se administra un test de aptitud a todos los aspirantes a puestos oficiales de una región. Se elige al azar una muestra de 50 aspirantes y estos son los resultados:

77	44	49	33	38	33	76	55	68	39
29	41	45	32	83	58	73	47	40	26
34	47	66	53	55	58	49	45	61	41
54	50	51	66	80	73	57	61	56	50
38	45	51	44	41	68	45	93	43	12

a. Construya una tabla de distribución de frecuencias.
b. Construya el histograma y el polígono de frecuencias.
c. Construya la ojiva.

23) En una empresa el personal se distribuye de acuerdo con su actividad desarrollada en la misma, como se indica a continuación:

Actividad	Porcentaje
Profesional	8%
Técnica	10%
Operario	70%
Ayudante	10%
Aseo	2%

Construya un gráfico circular para ilustrar la situación.

24) Los datos que se muestran a continuación, son los cargos (en dólares) por los servicios de electricidad, agua y gas durante el mes de julio del 2000 para una muestra de 50 apartamentos de 3 habitaciones en Caracas:

96	171	202	178	147	102	153	197	127	82
157	185	90	116	172	111	148	213	130	165
141	149	206	175	123	128	144	168	109	167
95	163	150	154	130	143	187	166	139	149
108	119	183	151	114	135	191	137	129	158

Elaborar:

a. Una tabla de distribución de frecuencias.
b. Un histograma y un polígono de frecuencias.
c. Un histograma porcentual.
d. Una ojiva.

Determinar:

a. El porcentaje de apartamentos cuyo gasto no llega a 139 dólares.
b. El porcentaje de apartamentos cuyo gasto es mayor o igual a 158 dólares.
c. El porcentaje de apartamentos cuyo gasto es al menos de 120 dólares, pero menor de 196 dólares.
d. El porcentaje de apartamentos con gastos menores de 135 dólares.
e. El porcentaje de apartamentos con gastos de 186 dólares o más.
f. El porcentaje de apartamentos con gastos entre 140 y 184 dólares.

25) Hallar la media aritmética, la mediana y la moda de los ejercicios (1), (3), (5), (6), (8), (10), (15) y (16)

26) Hallar la media aritmética, la mediana y la moda de los números: 3, 5, 2, 5, 9, 5, 2, 8, 6.
Sol: $\bar{x} = 5{,}10$ Med = 5 Mo = 5.

27) Las calificaciones de un estudiante en las 3 asignaturas del curso fueron 71, 78 y 89.

a) Si los pesos asignados a cada asignatura son 2, 4, y 5 respectivamente. ¿Cuál es el promedio adecuado para sus calificaciones?
Sol: $\bar{x} = 81{,}72$ puntos
b) ¿Cuál será el promedio si todos los pesos fuesen iguales?
Sol: $\bar{x} = 79{,}33$ puntos.

28) La siguiente tabla muestra una distribución de la carga máxima en toneladas que soportan ciertos cables producidos por una compañía. Determinar la media de la carga máxima:

Máximo de carga	Número de cables
[9,3-9,8)	2
[9,8-10,3)	5
[10,3-10,8)	12
[10,8-11,3)	17
[11,3-11,8)	14
[11,8-12,3)	6
[12,3-12,8)	3
[12,8-13,3)	1

Sol: $\bar{x} = 11{,}14$ Ton.

29) Una serie de números está formada por seis (6), siete (7), ocho (8), nueve (9), diez (10).
¿Cuál es su media aritmética?
Sol: $\bar{x} = 8{,}25$

30) Un investigador obtuvo las respuestas siguientes a una de las preguntas incluidas en una encuesta de evaluación: totalmente en contra, en contra, ligeramente en contra, un poco de acuerdo, de acuerdo, altamente de acuerdo, totalmente de acuerdo. ¿Cuál es la mediana?
Sol: un poco de acuerdo

31) Una empresa constructora tiene 2 secciones A y B. Las distribuciones de ingresos diarios de sus empleados son los siguientes:

Sección A

Ingresos ($)	Frecuencia
[80-100)	30
[100-120)	80
[120-140)	40
[140-160)	10
[160-180)	4
[180-200)	1

Sección B

Ingresos ($)	Frecuencia
[60-90)	10
[90-120)	20
[120-150)	50
[150-180)	20
[180-210)	15
[210-240)	10
[240-270)	4

¿Cuál es el ingreso de las dos secciones en conjunto?

Sol: $\bar{x}$ sección A = 115,57 $

$\bar{x}$ sección B = 148,02 $

Ingreso promedio pedido $\bar{x}_{media}$ = 129,80 $.

32) La distribución de frecuencia siguiente, representa los pesos en kilogramos de una muestra de paquetes transportados por una compañía aérea:

Pesos (Kg)	f_i
[10-11)	1
[11-12)	4
[12-13)	6
[13-14)	8
[14-15)	12
[15-16)	11
[16-17)	8
[17-18)	7
[18-19)	6
[19-20)	2

Calcule la media, mediana y la moda de la muestra.

Sol: $\bar{x}$ = 15,20 kg Me = 15,13 kg Mo = 14,80 kg.

33) Un profesor decide utilizar un promedio ponderado para obtener las calificaciones de los estudiantes que acuden al seminario que imparte. El promedio de tareas tendrá un valor de 20% de la calificación del estudiante, el examen semestral 25 %, el examen final 35%, el artículo de fin de semestre 10% y los exámenes parciales 10%. A partir de los datos siguientes, calcule el promedio final para cada estudiante del seminario.

Estudiante	Tareas	Parciales	Articulo	Ex. Semestral	Ex. Final
1	85	89	94	87	90
2	78	84	88	91	92
3	94	88	93	86	89

Sol: $\bar{x}_1$ = 88,55 puntos $\bar{x}_2$ = 87,75 puntos $\bar{x}_3$ = 89,55 puntos.

34) Para la siguiente distribución de frecuencias, determine la mediana.

Clase	f_i	F_i
[100-150)	12	12
[150-200)	14	26
[200-250)	27	53
[250-300)	58	53
[300-350)	72	183
[350-400)	63	246
[400-450)	36	282
[450-500)	18	300

Sol: Me = 327,08

35) ¿Cuáles son los valores modales para las siguientes distribuciones?

Color de cabello	frecuencia	Tipo de sangre	frecuencia
Negro	11	AB	4
Castaño	24	O	12
Pelirrojo	6	A	35
Rubio	18	B	16

Sol: Castaño y Tipo A

36) Para la siguiente tabla de distribución de frecuencias, calcule la media, la moda y la mediana del conjunto de datos.

Días	frecuencia
[0- 1)	2
[1- 2)	4
[2- 3)	6
[3- 4)	7
[4- 5)	5
[5- 6)	3
[6- 7)	1

Sol: $\bar{x}$ = 3,28 días Mo = 3,33 días Me = 3,28 días.

37) Hallar los cuartiles, los deciles, P_{80} y P_{14} e interpretar su significado de los ejercicios (1), (3), (5), (6), (8), (10), (15) y (16).

38) La tabla muestra una distribución de frecuencias de puntuaciones de un examen

Puntuación	Número de estudiantes
[30 – 40)	1
[40 – 50)	3
[50 – 60)	11
[60 – 70)	21
[70 – 80)	43
[80 – 90)	32
[90 – 100)	9

Hallar:

1. Los cuartiles de la distribución.
2. P_{12} e interprete su significado.
3. P_{68} e interprete su significado.
4. La puntuación más baja alcanzada por el 25% más alto del curso y la más alta alcanzada por el 20% más bajo del curso.
5. ¿A qué porcentaje corresponde una calificación de 58 puntos?.

Sol:

1. Q_1 = 67,14 puntos. Q_2 = 75,58 puntos. Q_3 = 83,43 puntos.
2. 59,45 puntos.
3. 80,81 puntos.
4. 83,43 y 64,28 puntos respectivamente.
5. 10,66%

39) Hallar la varianza y la desviación típica de los ejercicios (1), (3), (5), (6), (8), (10), (15) y (16).

40) Los datos adjuntos representan el promedio de millas por galón diario por 5 días para los carros A y B, en condiciones similares:

A	20	25	30	15	35
B	15	27	25	23	35

a) Encuentre la media y el rango de millas por galón para cada carro

b) ¿Cuál carro parece haber logrado un rendimiento más consistente, si la consistencia se determina examinando las varianzas?

Sol: a) Ambos carros tienen el mismo rango (R = 20)

Ambos carros tiene la misma media ($\bar{x}$ = 25 millas por galón)

b) El carro B es más consistente.

41) La tabla de frecuencias exhibe las edades de una muestra de 36 personas que asistieron a una película:

Años	f
8-13	2
14-19	7
20-25	13
26-31	5
32-37	9

Hallar:

a. La media Sol: 24,5 años
b. La varianza Sol: 53,48 años^2
c. La desviación típica Sol: 7,31 años.

42) En un examen final de estadística, la puntuación media de un grupo de 150 estudiantes fue de 78 y la desviación típica fue de 8 puntos. En álgebra, sin embargo, la media final del grupo fue de 73 y la desviación típica 7,6. ¿En qué asignatura hubo mayor dispersión absoluta y en cuál mayor dispersión relativa?

Sol: estadística y álgebra respectivamente.

43) Para comparar la precisión de 2 instrumentos de medición, un técnico de laboratorio estudia mediciones hechas con ambos instrumentos. El primero se usó recientemente para medir el diámetro de un rodamiento y las mediciones tuvieron una media de 4,92 mm. con una desviación estándar de 0,018 mm. El segundo se empleó hace poco para medir la longitud sin extender de un resorte y las mediciones tuvieron una media de 2,54 pulgadas con una desviación estándar de 0,012 pulgadas. ¿Cuál de los 2 instrumentos es relativamente más exacto?.
Sol: el primero

44) José Pérez es un directivo de una empresa de planificación financiera que asesora a quienes quieren establecer sus carteras de inversión personales. Hace poco José estaba interesado en las tasas de rendimiento que habían ofrecido dos fondos de inversión diferentes a lo largo de los 5 últimos años. FIVENEZ presentaba tasas de retorno a lo largo de ese período de 12, 10, 13,9 y 11%; mientras que Corporación Dinámica había producido 13, 12, 14, 10, y 6%. Un cliente se puso en contacto con el señor Pérez y expresó su interés por uno de estos fondos de inversión. ¿Cuál de ellos deberá elegir Pérez para su cliente?
Sol: FIVENEZ.

45) La señorita Disbier Araque utiliza 2 máquinas diferentes para fabricar productos de salida de papel destinadas a copiadoras Kodak. Los conductos de una muestra de la primera máquina medían 12,2 ; 11,9 ; 11,8 ; 12,1 ; 11,9 ; 12,4 ; 11,3 y 12,3 pulgadas. Los conductos hechos con la segunda máquina medían 12,2 ; 11,9 ; 11,5 ; 12,1 ; 12,2 ; 11,9 y 11,8 pulgadas. Disbier tiene que utilizar la máquina que produzca conductos de tamaños más uniformes. ¿Qué máquina deberá utilizar?
Sol: la segunda máquina.

46) La licenciada Denisse Villalba como directora de vuelo de United necesita información sobre la dispersión del número de pasajeros. Las decisiones en relación con los horarios y el tamaño más eficiente de los aviones dependen de la fluctuación de la carga de pasajeros. En base a esto y tomando en cuenta la siguiente tabla de frecuencias:

Pasajeros	f_i
[50-59)	2
[59-68)	5
[68-77)	14
[77-86)	18
[86-95)	7
[95-104)	4

¿Cuál debió ser la desviación calculada por Denisse?
Sol: 10,79 pasajeros.

47) En un estudio se anotó el número de palabras leídas en 15 segundos por un grupo de 120 sujetos que habían recibido previamente un adiestramiento y 120 individuos que no habían recibido dicha instrucción. Los resultados fueron los siguientes:

Número de palabras leídas	No instruidos	Instruidos
25	56	1
26	24	9
27	16	21

28	12	29
29	10	28
30	2	32

Compare la variabilidad en ambos grupos.
Sol: la variabilidad es parecida.

48) La señora Lorena de Ugueto es una agente de inversiones que encuentra 2 valores prometedores. El primero conduce a un rendimiento medio del 10% con una desviación típica del 1,2%; el segundo produce una tasa de rendimiento medio del 20% con una desviación típica del 5%. Con ayuda del CV como medida del riesgo, Lorena aconseja a su cliente más conservador que invierta en el primer valor. ¿Estaría usted de acuerdo?.
Sol: sí estaría de acuerdo.

49) La siguiente tabla representa la edad de los empleados que trabajan en cierta empresa:

Edad	N° de Empleados
[22-26)	12
[26-30)	29
[30-34)	27
[34-38)	19
[38-42)	16
[42-46)	10
[46-50)	7

Calcular:

a) Edad más frecuente
b) La edad que se encuentra justo en el 50% de la distribución.
c) La edad mínima del 40% entre los mayores
d) Porcentaje entre 28 y 40 años
e) Porcentaje entre $\bar{x} \pm s$
f) Calcule la curtosis e interprete
g) Calcule el coeficiente de asimetría e interprete.

Sol: a) 29,58 b) 32,81 c) 34,84 d) 57,09 e) 64,18% g) 0,47.

50) En un examen final de microeconomía, la puntuación media de 150 estudiantes fue 12,8 puntos y la desviación típica 2,3 puntos. En estadística el promedio fue 10,2 puntos y la desviación típica 1,6 puntos.

a. En qué materia hay mayor dispersión relativa?
b. ¿En qué materia destaca más un alumno que obtuvo 14 puntos en ambas?

Sol: a) En microeconomía, b) en estadística.

51) En cierta evaluación para optar por una beca, Yulimar Prato obtuvo una calificación de 310 puntos en habilidad verbal y 218 puntos en habilidad numérica. Los parámetros de c/u son:

Habilidad verbal: $\bar{x} = 245 \quad s^2 = 900$
Habilidad numérica: $\bar{x} = 150 \quad s = 24$

a. ¿En cuál de las dos pruebas obtuvo mejor calificación?
b. ¿En cuál de las dos pruebas el grupo es más homogéneo?

Sol: a) en Habilidad numérica b) en Habilidad verbal.

52) La siguiente tabla representa los resultados en la prueba de aptitud académica de un grupo de 1000 jóvenes que aspiran ingresar a cierta universidad:

Calificación	[300-350)	[350-400)	[400-450)	[450-500)	[500-550)	[550-600)
%Hi	6	28	45	63	95	100

Hallar.

a) Porcentaje de aspirantes cuya calificación es superior a 420 puntos pero inferior a 510
b) N° de estudiantes que obtuvieron 500 puntos o más
c) La mayor nota del 30% que obtuvo la nota más baja
d) Porcentaje que obtuvo más de 480 puntos
e) Coeficiente de asimetría de Pearson e interprete
f) La curtosis e interprete.

Sol: a) 34,6% b) 370 c) 405,88 d) 44,2% e) –0,69

53) El sueldo mensual, en miles de Bs., de los empleados de una empresa es como sigue:

Sueldo	[160-190)	[190-220)	[220-250)	[250-280)	[280-310)	[310-340)	[340-370)	[370-400)
fi	10	12	12	16	15	14	10	8

Elabore una tabla de distribución de frecuencias y calcule:

a. ¿Qué sueldo está en el centro de la distribución?
b. ¿Qué porcentaje gana Bs.270000 o menos?
c. ¿Qué intervalo abarca el 70% central?
d. Calcule la curtosis e interprete

Sol: a) 277,18 b) 46,04% c) [201,38 ; 350,35]

54) Se tiene una distribución simétrica y se conoce que el número de datos es 150, $f_3 = 30$ y $f_2 = f_1 + 4$. Construya la tabla de frecuencias y calcule el valor por debajo de la cual se halla el 60% de la muestra:

$l_i - l_{i+1}$
[3 – 6)
[6 – 9)
[9 – 12)
[12 – 15)
[15 – 18)

55) La media aritmética de dos números es igual a 6 y su desviación típica es igual a $2\sqrt{2}$. ¿Cuáles son esos números?

Sol: 4 y 8.

56) De la siguiente distribución de calificaciones en cierto examen calcular:

$l_i - l_{i+1}$	[6 – 8)	[8 –10)	[10 – 12)	[12 – 14)	[14 – 16)	[16 – 18)
F_i	3	10	25	38	42	45

Calcular:

a. Porcentaje de calificaciones entre $\bar{x} \pm 2s$.

Sol:
69,01%

b. El intervalo donde cae el 60% central. [9,71 ; 13,69]
c. Porcentaje de los que obtuvieron más de 15 puntos. 11,11%
d. La mayor nota del 25% que obtuvo la más baja. 10,16 puntos.
e. La menor nota del 25% que obtuvo la nota más alta. 13,34 puntos.
f. Coeficiente de asimetría e interprete. 0,064.
g. Calcule la curtosis, e interprete.

57) En las siguientes tablas se registran los sueldos quincenales (en miles) de 50 obreros de dos fábricas.

Fábrica A		Fábrica B	
Sueldo	$\%h_i$	Sueldo	$\%h_i$
[45 – 55)	18	[45 – 55)	12
[55 – 65)	24	[55 – 65)	28
[65 – 75)	26	[65 – 75)	30
[75 – 85)	20	[75 – 85)	22
[85 – 95)	12	[85 – 95)	8

a) ¿En cuál fábrica hay mayor dispersión relativa? Sol: A
b) Un obrero que gana Bs. 140000 mensuales. ¿Dónde estaría mejor remunerado con respecto a sus compañeros? Sol: B
c) ¿Cuál de las dos distribuciones es más simétrica? Sol: A

58) En la tabla se presenta una muestra de 25 empresas fabricantes de componentes de computadoras tomada de la base de datos de Stock Invester Pro.:

Empresa	**Precio por acción**	**Relación precio/valor en libros**	**Ganancias por acción ($ anuales)**
Amdahl	12,31	2,49	-2,49
Auspex Systems	11,00	2,22	0,85
Compaq Computer	65,50	6,84	2,01
Data General	35,94	4,25	1,15
Digi International	15,00	2,04	-0,89
Digital Equipment	43,00	1,92	-2,93
Pointe Technologies	14,25	3,47	0,80
Equitrac	16,25	2,38	0,76
Franklin Electronic	12,88	1,41	0,82
Gateway 2000	39,13	6,45	1,74
Hewlett-Packard	61,50	4,35	2,64
Ingram Micro	28,75	4,53	1,01
Maxwell Technologies	30,50	8,07	0,46
MicroAge	27,19	2,16	1,25
Micron Electronics	16,31	4,48	1,06
Network Computing	11,88	3,34	0,15
Pomeroy Computer	33,00	3,29	1,81
Sequent Computer	28,19	2,65	0,36
Silicon Graphics	27,44	3,01	-0,44
Southern Electronics	15,13	2,46	0,99
Stratus Computer	55,50	2,48	2,52

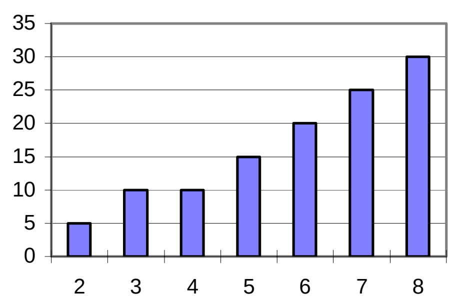
Sun Microsystems	48,00	7,50	1,67
Tandem Computers	34,25	3,61	1,02
Tech Data	38,94	3,80	1,50
Unisys	11,31	16,64	0,08

Elabore una tabla de distribución de frecuencias para los precios de las acciones y otra para los datos de las ganancias por acción.

59) Completar la siguiente tabla:

$[L_i - L_{i+1})$	x_i	f_i	hi	F_i	H_i
[0,10)		2	0,05	2	0,05
[10,20)					0,15
[20,30)					0,4
[30,40)		15			0,775
[40,50)					1

60) Construye la tabla de frecuencias a partir del siguiente gráfico de frecuencias acumuladas, sabiendo que tenemos una variable discreta



61) La distribución de los salarios mensuales en 1998 en la industria turística de nuestro país es reflejado en la siguiente tabla:

Salarios	f_i
0 – 50000	2145
50000 – 75000	1520
75000 – 100000	840
100000 – 115000	955
115000 – 135000	1110
135000 – 140000	2342
140000 – 150000	610
150000 – 200000	328
200000 o más	150

Calcular a partir de los datos:

a. El salario medio por trabajador. (marca del último intervalo = 300000).
b. El salario más frecuente.
c. Un salario tal que la mitad de los restantes sean inferiores a él.
d. El 35% de los trabajadores que más cobran. ¿Qué salario mínimo tienen?.

62) Los salarios mensuales de 4 individuos son Bs. 150000, 160000, 165000 y 200000 Hallar el salario medio. Ahora entra a trabajar una nueva persona en la empresa, percibiendo un salario de Bs. 500000 mensuales. ¿Se verá afectado el salario medio tras esta incorporación?. ¿Crees que la media es una medida de centralización adecuada en los dos casos?. En caso de que no lo sea, propón y calcula otra medida de centralización más adecuada.

63) En un taller de reparación de automóviles se recogen datos sobre los días de permanencia de los vehículos a reparar en él, y se obtiene:

Días de estancia	1	2	3	4	5	8	15
Nº de coches	23	12	7	10	3	2	1

a. Calcula el número medio de días de permanencia.
b. ¿Cuántos días como máximo permanecen en el taller el 75% de los automóviles, que menos permanecen en el taller?
c. Calcula la mediana y la moda

64) Tenemos dos variables X e Y con el mismo recorrido y media, siendo sus varianzas 4 y 9 respectivamente. ¿Para cual de las dos variables el valor de la media es más representativo?

65) Sea una variable con media 8 y desviación típica 0. ¿Qué se puede afirmar sobre el comportamiento de esta variable?.

66) La distribución de edades del Censo Electoral de Residentes al 1 de enero de 1999 para las comunidades autónomas de Aragón y Canarias, en tanto por cien es la siguiente:

Edades	**Aragón**	**Canarias**
16–18	3,54	4,35
18–30	21,56	29,99
30–50	31,63	35,21
50–70	28,14	21,97
70–90	15,13	8,48

a) Representa sobre los mismos ejes de coordenadas los histogramas de la distribución de la edad para las dos comunidades autónomas (emplea distinto trazo o distintos colores). ¿Que conclusiones obtienes a la vista de los histogramas?
b) A partir del gráfico determine qué comunidad tiene mayor variabilidad en la distribución de su edad.

Ejercicios Complementarios

1) Se les pidió a un grupo de estudiantes de la Universidad que hiciera una valoración global de la vida en una residencia universitaria. El dueño de la residencia quisiera alojar sólo a hombres porque piensa que éstos son mucho menos exigentes que las mujeres. Los resultados promedios obtenidos en una encuesta de valoración de la residencia entre 1(baja) y 10(alta), fueron los siguientes.
Hombres: 6,8 ; 9,1 ; 8,0 8,8 ; 8,8 ; 9,59 ; 7,90 ; 8,0 ; 8,0 ; 8,3 ; 10,0 ; 6,0 ; 6,5
Mujeres: 7,0 ; 7,1 ; 7,4 ; 7,3 ; 7,0 ; 7,5 ; 7,2 ; 7,15 ; 6,9 ; 6,8 ; 7,0 ; 7,3

Realice un análisis descriptivo de los datos que incluya gráficos y las medidas descriptivas adecuadas para investigar si realmente el dueño tiene razón en lo planteado.

2) Complete las líneas en blanco que aparecen a continuación.

a) Las medidas de tendencia central que siempre existen son _________ y _________.
b) Cuando existen datos extremos no es adecuado el empleo de la ___________ como medida de tendencia central.
c) La medida de variabilidad que es adecuada calcular cuando las medias de los grupos difieren es ______________.
d) Cuando aproximadamente el 65% de los datos está en el intervalo [x-s; x+s] la distribución de los datos es:_________________.
e) La clase ________es aquella dónde la frecuencia es mayor.
f) El rango intercuantil se calcula como_______________.
g) Los _______________ dividen a la distribución en cuatro partes iguales.
h) La varianza indica la distancia promedio de cualquier observación del conjunto de datos con respecto a _______________.
i) La diferencia entre el valor más alto de un conjunto de datos y el mínimo se conoce como ____________.

3) Los siguientes datos corresponden a las ventas en miles de dólares de una empresa en un período de 3 años

12, 27, 14, 27, 15, 19, 26, 25, 25, 16, 16,18, 18, 17, 19, 20, 23, 18
24, 23, 24, 19, 23, 22, 16, 16, 21, 19, 20, 22, 21, 20, 22, 21,16, 19

a) Construya la tabla de distribución de frecuencias.
b) ¿Cuál es la venta promedio de esta empresa?
c) Analice si la distribución es simétrica.
d) Se considera que una empresa es rentable en un período si el 75% de las ventas del período es superior a los 21000 dólares. De acuerdo a este criterio ¿es la empresa rentable?.
e) ¿Cuál es el valor mayor de la venta que hace la empresa entre el 20% de aquellos meses que la empresa menos vende?

4) Lo siguientes datos corresponden al número de turistas extranjeros que visitan una zona de Venezuela en 2 etapas del año:
Etapa 1: Enero - Julio:
108, 112, 94, 144, 162, 162, 76, 102, 11, 79, 129, 95, 114,121.
Etapa 2: Agosto - Diciembre
30, 41, 41, 40, 43, 25, 32, 22, 27, 64, 33, 41.

Según una agencia de viajes la variabilidad de los datos debe ser mayor en el primer período. ¿Se corresponde esta hipótesis con lo observado en las muestras?. Realice los cálculos que usted considere necesario para dar respuesta a esta interrogante, y justifique respuesta.

5) Complete las siguientes afirmaciones:
a. Un _______________ es un gráfico conformado por una serie de rectángulos cuyas alturas dependen de la frecuencia absoluta o relativa de cada clase.

c. Una _______________ es una colección de todos los elementos de un grupo. Un subconjunto de todos los elementos de un grupo es ____________.

6) El dueño de un restaurante está interesado en estudiar los patrones de consumo de sus clientes. Con este fin tomó una muestra de 20 clientes para los que registró el tipo de plato que ordenó, si ordenó o no postre y el monto en bolívares del consuno realizado por cada cliente. Los datos se presentan en la siguiente tabla:

Tipo de Plato	**Postre**	**Monto Consumo**
Carne	No	7721
Pollo	Sí	7130
Pollo	No	6274
Carne	Sí	7836
Pescado	No	7382
Pollo	Sí	7371
Carne	No	4407
Pollo	Sí	7176
Pescado	Sí	8074
Carne	No	6947
Pescado	Sí	7699
Pollo	Sí	6507
Carne	Sí	9321
Pescado	Sí	7532
Pollo	No	6893
Carne	No	6392
Pescado	Sí	7454
Carne	Sí	7741
Pollo	Sí	7300
Carne	Sí	7923

a) El dueño está interesado en saber cómo está relacionado el monto del consumo con el tipo de plato, para saber a qué tipo de plato le daría prioridad en la preparación. Ayúdelo en su selección, tomando en cuenta todas las características de los datos y construyendo un gráfico que facilite la comprensión. Escriba sus conclusiones.

b) ¿Los datos apoyan la siguiente afirmación: "Los clientes que consumen carnes blancas en su mayoría piden postre"?

c) Si un cliente cualquiera desea ir al restaurante a comer pescado, ¿entre qué valores oscilaría el monto de su consumo?

d) ¿Cuánto sería el monto del consumo mínimo del 10% de los clientes que más gastan en el restaurante?

7) Una gran compañía llevó a cabo un estudio para ubicar las variables que pudieran determinar el sueldo de un egresado universitario dos años después de haberse graduado como Técnico Superior Universitario en un área Administrativa. Los datos recogidos se presentan en la siguiente tabla:
(La columna del sueldo es en cientos de miles de bolívares.)

	Edad	**Sexo**	**E. Civil**	**Inglés**	**Sueldo**
1	24	F	C	A	6,75
2	25	M	C	M	6,90
3	26	M	S	B	6,90
4	27	F	C	B	6,80
5	27	M	D	A	7,10
6	27	F	C	M	6,50
7	27	M	S	A	7,25
8	25	F	C	B	6,80
9	23	M	S	B	6,75
10	24	M	S	B	6,80
11	26	F	C	M	6,75
12	29	F	D	M	7,00
13	25	M	C	A	7,15
14	31	F	D	A	7,50
15	26	M	S	B	6,20
16	24	F	D	M	7,40
17	26	F	C	B	6,70
18	28	F	S	M	6,95
19	25	M	C	B	6,95
20	29	M	C	M	7,10

a) Utilice la técnica de estadística descriptiva más apropiada para analizar cada variable individualmente. Interprete lo obtenido.
b) Realice diagramas de cajas que le ayuden a visualizar como influye cada una de las variables en el sueldo que gana el individuo.
c) Como futuro Técnico Superior en el área Administrativa, ¿cuál seria la(s) características que usted debería tomar en consideración para obtener el sueldo al que usted aspiraría al egresar?

8) Durante una epidemia de gripe, los tiempos de espera en cierto centro de salud fueron más largo de lo habitual. La siguiente tabla resume la distribución de los tiempos de espera para una muestra de 20 pacientes que visitaron el centro de salud durante este periodo,

Tiempo de Espera (horas)	[0 - 1)	[1 - 2)	[2 - 3)	[3 - 4]
N° de Pacientes	6	9	4	1

a) Estime el tiempo medio de espera de los pacientes y la variabilidad del mismo.
b) Un nuevo paciente que llega al centro de salud plantea que él esperará para ser atendido si al menos la mitad de los pacientes es atendido en 1,5 horas. ¿Esperará el paciente para ser atendido? ¿Por qué?
c) Calcule un valor aproximado del tiempo de mayor frecuencia para ser atendido un paciente.
d) El centro de salud ha decidido que al 20% de los pacientes que más tardan en ser atendidos, los remitirá a una consulta adicional que ha habilitado. ¿Cuánto tiempo tiene que esperar como mínimo un paciente para ser enviado a esta consulta especial?

9) Clasifica las siguientes variables como cualitativas o cuantitativas, y a estas últimas como continuas o discretas:

a) Carreras que se estudian en la U.S.B.
b) Nº de cartas que se escriben en un mes
c) Número de calzado
d) Precio de un producto.
e) Marcas de cerveza
f) Nº de empleados de una empresa
g) Altura
h) Temperatura de un enfermo

10) La siguiente tabla muestra los coeficientes de inteligencia (C.I) de 480 niños de una escuela elemental.

Clases	[70-78)	[78-86)	[86-94)	[94-102)	[102-110)	[110-118)	[118-126)	[126-134]
f_i	4	16	45	85	54	27	11	2

a) Si una madre afirma que exactamente la mitad de los niños del colegio tienen un C.I. superior al de su hijo. ¿Qué coeficiente de inteligencia tiene su hijo?.
b) Supongamos que se quieren hacer estudios sobre el proceso de aprendizaje de los niños con mayor C.I., pero que el psicólogo solo puede atender al 15% de los niños del colegio. ¿Qué C.I. deberá tener un niño como mínimo para ser considerado dentro de ese grupo de elegidos?
c) ¿Qué porcentaje de niños tiene un coeficiente de inteligencia menor a 80?
d) ¿Es la distribución de estos datos simétrica?. Justifique su respuesta.
e) Represente en un gráfico la información acerca de los coeficientes de inteligencia.

11) Supongamos que dos empresas desean repartir beneficios entre sus cuatro principales accionistas, y que el reparto se realiza de la siguiente forma (en miles de bolívares)

	Empresa A	**Empresa B**
1r accionista	100	1200
2º accionista	500	1300
3r accionista	300	1400
4º accionista	100	1100

¿Cuál de los dos repartos el más equitativo?. Justifica el resultado en base al análisis estadístico del reparto.

12) En lo que va de año cierta tienda deportiva ubicada en un prestigiado centro comercial de la ciudad capital ha vendido la última línea de zapatos deportivos Nike, Skechers, Addidas y Reebok. A continuación se presentan los datos sobre 50 ventas de zapatos deportivos en dicha tienda:

Nike	Nike	Nike	Skechers	Nike	Addidas	Nike	Addidas	Skechers	Skechers
Skechers	Skechers	Addidas	Nike	Skechers	Nike	Nike	Nike	Addidas	Nike
Nike	Addidas	Addidas	Reebok	Nike	Nike	Nike	Skechers	Nike	Skechers
Nike	Nike	Skechers	Skechers	Nike	Nike	Addidas	Skechers	Skechers	Skechers
Skechers	Nike	Nike	Addidas	Nike	Skechers	Nike	Skechers	Nike	Nike

Utilizando las técnicas estadística descriptiva (tabla, gráfico, etc.), realice un breve informe con relación a lo que sugieren los datos sobre las ventas de zapatos deportivos en dicha tienda.

13) Durante los 9 primeros meses del 2004, los destinos turísticos más visitados en el Caribe y Europa fueron Orlando, Portugal, Canaima, Margarita y Aruba, según la empresa AEROTUR. A continuación se presentan los datos sobre los 50 destinos turísticos mas visitados en el Caribe y Europa:

Orlando	Portugal	Canaima	Aruba	Aruba
Portugal	Portugal	Aruba	Portugal	Aruba
Canaima	Margarita	Canaima	Portugal	Aruba
Portugal	Portugal	Aruba	Aruba	Margarita
Portugal	Portugal	Margarita	Margarita	Aruba
Margarita	Aruba	Aruba	Portugal	Portugal
Margarita	Portugal	Margarita	Aruba	Margarita
Canaima	Margarita	Canaima	Aruba	Portugal
Canaima	Canaima	Portugal	Aruba	Margarita
Portugal	Orlando	Margarita	Canaima	Portugal

Utilizando técnicas de estadísticas descriptiva (tabla, grafico, etc.), realice un breve informe con relación a lo que sugieren los datos sobre los destinos turísticos mas visitados.

Ejercicios (Probabilidades)

1) Una carta se extrae aleatoriamente de una baraja de 52. Encontrar la probabilidad de que sea:

a. Un as Sol: 1/13
b. Diez de corazones Sol: 1/52
c. Un 3 de tréboles o un 6 de diamantes Sol: 1/26
d. Un corazón Sol: 1/ 4
e. Un 10 o una pica Sol: 4/13
f. Ni un 4 ni un trébol Sol:9/13

2) Una esfera se extrae aleatoriamente de una caja que contiene 6 esferas rojas, 4 blancas y 5 azules. Determinar la probabilidad de que sea:

a. Rojo Sol: 2/5
b. Blanca Sol: 4/15
c. Roja o blanca Sol: 2/3

3) Una caja contiene 10 esferas rojas, 30 blancas, 20 azules y 15 naranjas, si se extrae aleatoriamente una de ellas determinar:

a. La probabilidad de que sea naranja o roja Sol: 1/3
b. La probabilidad de que sea blanca, roja o azul Sol: 4/5

4) Con los mismos datos del problema anterior, suponer que se extraen 2 esferas sucesivamente de la caja y que hay remplazamiento de la esfera extraída después de cada extracción para determinar:

a. La probabilidad de que ambas sean blancas Sol: 4/25
b. La probabilidad de que la 1^{era} sea roja y la 2da sea blanca Sol: 4/75
c. La probabilidad de que sean rojas o blancas o de ambos colores (rojas y blancas) Sol: 64/225

5) Se hacen 2 extracciones de una baraja de 52 cartas. Hallar la probabilidad de que las 2 cartas extraídas sean ases, siendo las extracciones sin remplazamiento.
Sol: 1/221

6) Se extraen sucesivamente 3 esferas de una caja que contiene 6 esferas rojas, 4 blancas y 5 azules. Hallar la probabilidad de que sean extraídas en el orden roja, blanca y azul, si las extracciones son:

a. Con remplazamiento Sol: 8/225
b. Sin remplazamiento Sol: 4/91

7) Una caja A contiene 3 pelotas rojas y 2 azules en tanto que una caja B contiene 2 pelotas rojas y 8 azules. Daisy lanza una moneda honrada y si se obtiene cara saca una pelota de la caja A; si se obtiene sello se saca de la caja B. Hallar la probabilidad de que Daisy saque una pelota roja.
Sol:2/5

8) Suponga ahora que Daisy no revela si resulta cara o sello (de tal forma que la caja de la cual sacó la pelota no se revela), pero revela que se sacó una pelota roja. ¿Cuál es la probabilidad de que se escogiera la caja A? (es decir, que el resultado de la moneda sea cara)

Sol: 3/4

9) Suponiendo que se tiene 3 cajas de las cuales se sabe que la caja I tiene 2 pelotas blancas y 3 negras; la caja II tiene 4 blancas y 1 negra y la caja III tiene 3 blancas y 4 negras. Se selecciona una caja aleatoriamente y una pelota extraída aleatoriamente es blanca. Hallar la probabilidad de haber escogido la caja I.
Sol: 14/57

10) La compañía Microtel desea mejorar la resistencia de las computadoras personales que construye, con respecto a fallas en la unidad de disco y el teclado. En la actualidad, el diseño de sus computadoras es tal que las fallas de la unidad de disco significan un tercio de las fallas del teclado. La probabilidad de que se presente una falla conjunta en la unidad de disco y en el teclado es de 0,05.
a. Si la computadora es 80% resistente a fallas en la unidad de disco o en el teclado. ¿Qué tan baja debe ser la probabilidad de que se presente una falla en la unidad de disco?
Sol: 0,0625
b. Si el teclado mejoró de tal modo que sólo falla el doble de veces que la unidad de disco (y la probabilidad de falla conjunta sigue siendo de 0,05). ¿La probabilidad de que la unidad de disco del inciso (a) producirá una resistencia a fallas en la unidad de disco duro, en el teclado, o en ambos, es mayor o menor que 90%?
Sol: menor (86,25%)

11) Susana Rivero es una consultora de una compañía publicitaria que lanzó recientemente una campaña para un nuevo restaurante. Susana acaba de instalar 4 anuncios panorámicos en la carretera a la entrada de la ciudad, y sabe, por su experiencia, la probabilidad de que cada anuncio sea visto por un conductor escogido aleatoriamente. La probabilidad de que el 1° anuncio sea visto por un conductor es de 0,75. La probabilidad de que el 2° sea visto es de 0,82 ; la probabilidad para el 3° es de 0,87 y la del 4° es de 0,9. Suponiendo que el evento, consistente en que un conductor vea cualquiera de los anuncios, es independiente de si ha visto o no los demás. Calcular la probabilidad de que:

a. Los 4 anuncios sean vistos por un conductor escogido aleatoriamente	Sol: 0,481545
b. El 1° y el 4° anuncio sean vistos, sin que el 2° y el 3° sean notados	Sol: 0,015795
c. Exactamente uno delos anuncios sea visto	Sol: 0,0316
d. Ninguno de los anuncios sea visto	Sol. 0,000585

12) María Campos, gerente del departamento de crédito de un banco, sabe que la compañía utiliza 3 métodos para conminar a pagar a las personas con cuentas morosas. De los datos que se tiene registrados, ella sabe que 70% de los deudores son visitados personalmente, 20% se le sugiere que paguen vía telefónica y al restante 10% se le envía una carta. Las probabilidades de recibir alguna cantidad de dinero debido a los pagos de una cuenta con estos 3 métodos son 0,75 0,60; y 0,65 respectivamente. La señorita Campos acaba de recibir el pago de una de las cuentas vencidas. Calcular la probabilidad de que la petición de pago se haya hecho:

a. Personalmente	Sol: 0,739
b. Por teléfono	Sol: 0,169
c. Por correo	Sol: 0,092

13) Una empresa compra cierto tipo de pieza que es suministrada por 3 proveedores: el 45% de las piezas son compradas al 1er proveedor resultando defectuoso el 1%, el 2do proveedor suministra 30% de las piezas y de ellas es defectuoso el 2%. Las restantes piezas provienen del 3er proveedor, siendo defectuoso el 3% de las mismas.
En un control de recepción de artículos se selecciona una pieza al azar y es defectuosa. Calcular la probabilidad de que la haya suministrado el 2do proveedor.
Sol: 0,3333

14) En un laboratorio se experimenta sobre una enfermedad que puede estar producida por 3 virus: A, B, C.
Hay 3 tubos de ensayo con el virus A, 2 tubos de ensayo con el virus B y 5 tubos de ensayo con el virus C. La probabilidad de que el virus A produzca la enfermedad es de $1/3$, que la produzca B es de $2/3$ y que la produzca el virus C es de $1/7$.
Se inocula un virus a un animal y contrae la enfermedad. ¿Cuál es la probabilidad de que el virus inoculado sea el C?.
Sol: 0,2343

15) Un conjunto normal de esferas de billar consta de 15 esferas numeradas del 1 al 15. Pegley Woodhull, el famoso jugador de billar ciego, está interviniendo en el juego conocido como bola 8, en el que esta esfera debe meterse de última.
Se le permiten tocar las esferas para determinar su posición antes de jugar, pero no sabe qué número tienen. Todos los tiros que hace Woodhull son buenos.

a. ¿Cuál es la probabilidad de que meta en la buchaca la esfera 8 en su 1er tiro, perdiendo así el juego?. Sol: $1/15$
b. ¿Cuál es la probabilidad de que la esfera 8 sea una de las tres primeras que meta?.Sol: $1/5$
c. ¿Cuál es la probabilidad de que gane el juego, esto es, que la esfera 8 sea la última en entrar a la buchaca?. Sol: $1/15$

16) Un transportista de productos tiene 10000 cajas de plátanos que vienen de Ecuador y de Honduras. Una inspección de la carga ha arrojado la información siguiente:

# de cajas con / # de cajas	Fruta dañada	Fruta muy madura
6000 (E)	200	840
4000 (H)	365	295

a. ¿Cuál es la probabilidad de que una caja seleccionada al azar contenga fruta dañada?
b. ¿Cuál es la probabilidad de que una caja seleccionada al azar sea ecuatoriana o de Honduras?.
c. Dado que una caja seleccionada al azar contiene fruta muy madura, ¿cuál es la probabilidad de que venga de Honduras?
d. Si tener fruta dañada y fruta muy madura son eventos mutuamente excluyentes. ¿Cuál es la probabilidad de que una caja contenga fruta dañada o fruta muy madura?.

Sol: 0,0565 ; 1 ; 0,2599 ; 0,17 respectivamente.

17) Un terapeuta físico que trabaja en la universidad sabe que el equipo de fútbol jugará 40% de sus juegos en campos con pasto artificial en la presente temporada. También sabe que las posibilidades de que un jugador de fútbol sufra una lesión en la rodilla son 50% más altas si juega en pasto artificial en lugar de hacerlo en pasto natural.
Si la probabilidad de que un jugador sufra una lesión en la rodilla mientras juega en pasto artificial es de 0,42.

a. ¿Cuál es probabilidad de que un jugador elegido aleatoriamente sufra una lesión en la rodilla?. Sol: 0,336
b. ¿Cuál es la probabilidad de que un jugador elegido aleatoriamente con lesión en la rodilla haya sufrido ésta mientras jugaba en un campo con pasto natural? Sol: 0,5

18) Si una moneda equilibrada se lanza al aire dos veces. ¿Cuál es la probabilidad de que caiga al menos una cara?
Sol: 0,75

19) Un bolsa contiene 6 metras azules, 2 rojas y 2 verdes. Si se selecciona una al azar, ¿cuál es la probabilidad de que sea:

a. Roja Sol: 1/5
b. Blanca Sol: 0
c. Verde Sol: 1/5
d. Azul Sol: 3/5

20) Hallar la probabilidad de sacar al azar una bola que no sea roja de una caja que contiene 3 bolas blancas, 2 rojas y 5 verdes.
Sol: 4/5

21) En una reunión se encuentran 10 personas de las cuales tres son educadores, 5 son contadores y dos economistas. Suponga que las personas tiene una sola profesión. Determinar la probabilidad de que una persona elegida al azar sea economista o contador.
Sol: 0,70

22) Calcular la probabilidad de que al lanzar un dado se obtenga un número que sea múltiplo de dos o tres.
Sol: 4/6

23) Si se lanzan dos dados, encontrar la probabilidad de obtener un 5 en el primero y 3 en el segundo.
Sol: 1/36

24) En cierto estado, el 25% de los automóviles emiten una excesiva cantidad de contaminantes. Si la probabilidad de que un automóvil que emite excesiva cantidad de contaminantes no pase la prueba de revisión vehicular es de 0,99 y la probabilidad de que un automóvil que no emite cantidad excesiva de contaminantes repruebe es de 0,17. ¿Cuál es la probabilidad de que un automóvil que no pase la prueba en realidad provenga de los que emiten cantidades excesivas de contaminantes?
Sol: 0,66

25) En una planta de electrónica se sabe, por experiencias pasadas, que la probabilidad de que un nuevo trabajador que ha asistido al Programa de Capacitación de la compañía cumpla con la cuota de producción es del 84%, y que la probabilidad de que un nuevo empleado cumpla con su cuota de producción sin haber asistido al Programa de Capacitación es de 0,49. Si el 70% de los trabajadores que ingresan como nuevos empleados asisten al Programa. ¿Cuál es la probabilidad de que un nuevo trabajador que cumpla con su cuota de producción, haya asistido al Programa de Capacitación?
Sol: 0,80

26) Una compañía de ventas por correo tiene tres empleados de almacén denominados U, V y W quienes toman productos de la bodega y los ensamblan para la subsiguiente verificación y empaquetado. U comete un error en un pedido (toma un producto equivocado o la cantidad equivocada del producto) una de cada 100 veces, V comete un error en un pedido 5 veces de cada 100 y W se equivoca tres de cada 100. Si U, V y W cubren respectivamente el 30%, el 40% y el 30% de todos los pedidos.
¿Cuál es la probabilidad de que si se encuentra un error en un pedido, éste haya sido cometido por V?
Sol: 0,625

27) De una caja que contiene pelotas numeradas del 1 al 6 se eligen dos, de forma consecutiva, sin reemplazo. Hallar:

a. La probabilidad de que en la segunda extracción salga un 5. Sol: $1/6$
b. La probabilidad de que salga un 2 en la 1^{ra} extracción y un 5 en la 2^{da} Sol: $1/30$
c. Supongamos ahora, que después de anotar el resultado de la primera extracción, se devuelve la pelota a la caja y se saca nuevamente una pelota. Hallar la probabilidad de los dos casos anteriores. Sol: $1/36$

28) De cuarenta cartas se elige primero una carta y a continuación se toma una segunda carta sin haber devuelto la primera al mazo. ¿Cuál es la probabilidad de que ambas sean oro? y ¿cuál es la probabilidad de que salgan dos ases?
Sol: 0,0576 y 0,0076 respectivamente

29) Supóngase que nos interesa la conclusión de la obra de construcción de una autopista, la cual puede demorarse por una huelga. Además suponga que las probabilidades son de 0,60 de que habrá una huelga, del 85% de que el trabajo se concluirá a tiempo si no hay huelga y de 0,35 de que el trabajo se terminará a tiempo si ocurre la huelga; si nos encontramos con que la obra se terminó a tiempo. ¿Cuál es la probabilidad de que pese a ello hubiese estallado una huelga?
Sol: 0,3818

30) Suponga que la probabilidad de que un hombre de 25 años quiera celebrar su quincuagésimo cumpleaños es de 74,2% y la probabilidad de que una mujer de 22 años viva hasta su cuadragésimo séptimo cumpleaños sea de 0,902. Si ambas personas se casan este año, ¿cuál es la probabilidad de que esa pareja viva para celebrar sus bodas de plata?
Sol: 0,6693

31) Una caja contiene 8 esferas rojas, 3 blancas y 9 azules. Si se extraen 3 esferas aleatoriamente sin reemplazo, determinar la probabilidad de que se extraiga una de cada color.
Sol: 18/95

32) Un especialista en alergias afirma que el 50% de los pacientes que examina son alérgicos a algún tipo de hierba.

a. ¿Cuál es la probabilidad de que exactamente 3 de sus siguientes 4 pacientes sean alérgicos?
b. ¿Cuál es la probabilidad de que ninguno de sus 4 pacientes sea alérgico?

Sol: 1/4 y 1/16 respectivamente

33) Se lanza una moneda con una probabilidad de 2/3 que el resultado sea cara; si aparece una cara, se extrae una pelota, aleatoriamente, de una caja que contiene 2 pelotas rojas y 3 verdes. Si el resultado es sello se extrae una pelota, de otra caja, que contiene 2 pelotas rojas y 2 verdes. ¿Cuál es la probabilidad de extraer una pelota roja?
Sol: 13/30

34) Se tienen 15 piezas de las cuales 5 son defectuosas. Si se seleccionan 3 piezas al azar, calcular la probabilidad de encontrar:

a. Ninguna defectuosa.
b. Al menos una defectuosa.

Sol: 0,263 ; 0,737 respectivamente.

35) Se seleccionan 2 fichas de dominó. ¿Cuál es la probabilidad de que queden encadenadas?
Sol: 0,389

36) En un juego de dominó. ¿Cuál es la probabilidad de sacar 4 seis en una mano?
Sol: 0,039

37) Se lanza un dado 5 veces. ¿Cuál es la probabilidad de que salga un 4, un 5 o un 6?
Sol: 15/96

38) En un estante hay 7 libros de historia y 3 de matemáticas. De los libros de historia, tres están empastados de amarillo y el resto de rojo; mientras que de los libros de matemáticas, uno está empastado en amarillo y dos en rojo. Suponiendo que del estante se elige un libro al azar, ¿cuál es la probabilidad de que sea matemáticas y rojo?
Sol: 0,20

39) Hindy Poon se encuentra preparando un informe para la empresa en que trabaja, la cual es Aeropostal, que a su vez será entregado al Departamento Federal de Aviación de Venezuela. El informe deberá ser aprobado, en primer lugar, por el responsable del grupo del cual Hindy es integrante, luego por el jefe de su departamento y después por el jefe de división, en ese orden. Hindy sabe que los tres directores actúan de manera independiente, además sabe que su responsable de grupo aprueba el 85% de sus trabajos, el jefe de su departamento rechaza dos de cada diez informes preparados por Hindy, y el jefe de división aprueba el 82% de los trabajos de Hindy. Dado esto calcular:

a. Probabilidad de que la primera versión del informe sea enviada al Departamento Federal de Aviación de Venezuela. Sol: 0,5576
b. Probabilidad de que la primera versión del informe sea aprobada por su responsable de grupo y por su jefe de departamento, pero no por el jefe de su división. Sol: 0,1224

40) De un urna que contiene pelotas numeradas del 1 al 6 se eligen dos, de forma consecutiva, sin reemplazo. Hallar:
a. Probabilidad de que en la segunda extracción salga un 5 Sol: 1/6
b. ¿Cuál es la probabilidad de que salga un 2 en la primera extracción y un 5 en la segunda? Sol: 1/30

41) ¿Cuál es la probabilidad de sacar menos de cinco puntos como resultado del lanzamiento de dos dados?
Sol: 1/6

42) Una caja contiene 10 tornillos, de los cuales 3 están defectuosos. Si se extraen 2 al azar, hallar la probabilidad de que los 2 estén buenos.
Sol: 0,4666

43) Una caja contiene 6 fichas numeradas del 1 al 6, si se extraen una por una sin reemplazo. ¿Cuál es la probabilidad de sacarlas en forma ordenada del 1 al 6?
Sol: 0,00138

44) En una encuesta entre alumnos de maestría en administración se obtuvieron los datos siguientes acerca de "el principal motivo del alumno para solicitar su ingreso a la escuela donde está matriculado".

Motivo / Tipo est.	Calidad de la escuela	Costo o comodidad	Otros	Totales
Tiempo completo	421	393	76	890
Tiempo parcial	400	593	46	1039
Totales	821	986	122	1929

a. Si un alumno es de tiempo completo. ¿Cuál es la probabilidad de que la calidad de la institución sea el principal motivo para elegir su escuela?. Sol: 0,47303
b. Si un alumno es de tiempo parcial. ¿Cuál es la probabilidad de que la calidad de la escuela sea el motivo para elegirla? Sol: 0,3849
c. Sea A el evento en que el alumno es de tiempo completo y sea B el evento que el alumno menciona que la calidad de la escuela es el 1er motivo de su solicitud. ¿Son independientes los eventos A y B?. Justifique se respuesta. Sol: no

45) En un mismo día, el equipo profesional de baloncesto de una ciudad juega como local, mientras que el equipo profesional de hockey de la misma ciudad juega como visitante. Según las probabilidades publicadas en la revista *Chance,* para los deportes profesionales, un equipo profesional de baloncesto tiene una probabilidad de 0,641 de ganar un juego cada día como local,

y uno de hockey tiene una probabilidad de 0,462 de ganar un juego cuando es visitante. Históricamente, cuando ambos equipos juegan en el mismo día, la probabilidad de que la nota principal de los noticieros del día siguiente se refiera al juego de baloncesto es de 60% y la del juego de hockey es de 40%.
Suponga que a la mañana siguiente de un día con este tipo de encuentros el encabezado de la sección deportiva es “Ganamos”. ¿Cuál es la probabilidad de que la nota sea acerca del equipo de baloncesto?.
Sol: 0,675

46) Antes de que un libro sea lanzado al mercado se recogen las reacciones de un grupo de personas a las que se les permite leer el libro previamente. Posteriormente a las ventas del libro se les asigna el calificativo de altas, moderadas o bajas de acuerdo a las noemas del mercado. Los resultados se muestran en la siguiente tabla:

Reacciones / Ventas	Favorables	Neutral	Desfavorables
Altas	173	101	61
Moderadas	88	211	70
Bajas	42	113	141

a) ¿Cuál es la probabilidad de que las ventas sean altas?
b) ¿Cuál es la probabilidad de que las reacciones sean favorables?
c) Si la reacción del grupo es favorable?. ¿Cuál es la probabilidad de que las ventas sean altas?
d) Si las ventas son bajas ¿Cual es la probabilidad de que las opiniones hayan sido desfavorables?
e) ¿Cuál es la probabilidad de que las opiniones sean favorables y las ventas sean altas?
f) ¿Cuál es la probabilidad de que las ventas sean favorables o desfavorables?. ¿Son esos sucesos mutuamente excluyentes? Justifique
g) ¿Son los sucesos “Opiniones desfavorables” y “Ventas Bajas” independientes? Justifique.

47) En un estudio realizado para un supermercado se clasifican los clientes en aquellos que visitan el establecimiento de una manera frecuente u ocasional y de acuerdo a la frecuencia en que adquieren cierto alimento. En la siguiente tabla se presentan las proporciones correspondientes a cada uno de los grupos.

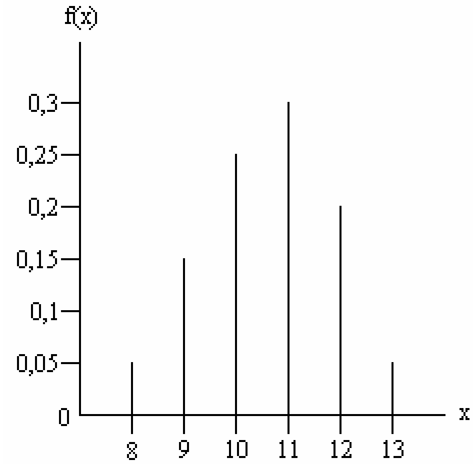
Compra de productos / Frecuencia en las visitas	Regular	Ocasional	Nunca
Frecuentes	0,12	0,48	0,19
No Frecuentes	0,07	0,06	0,08

a) ¿Cual es la probabilidad de que un cliente visite frecuentemente el supermercado y compre regularmente el producto alimenticio?
b) ¿Cuál es la probabilidad de que un cliente que nunca compra el producto visite el supermercado frecuentemente?
c) ¿Son los sucesos “Nunca compra productos alimenticios” y “Visita el mercado frecuentemente” independientes?. Justifique.
d) ¿Cuál es la probabilidad de que un cliente realice compras ocasionales?
e) ¿Cuál es la probabilidad de que un cliente no realice nunca compras del producto?

f) ¿Cuál es la probabilidad de que un cliente visite el establecimiento frecuentemente o compre el producto regularmente?

Ejercicios (Variable Aleatoria Discreta)

1) A partir de la gráfica siguiente, de una función de probabilidad:



a. Construya la tabla de la función de probabilidad.
b. Encuentre el valor esperado de la variable aleatoria. Sol: E(**X**) = 10,6

2) La única información disponible que tiene usted con respecto a la distribución de probabilidad de un conjunto de resultados es la siguiente lista de frecuencias:

x	0	15	30	45	60	75
f_i	25	125	75	175	75	25

a. Construya la función de probabilidad para el conjunto de datos.
b. Encuentre el valor esperado. Sol: E(**X**) = 36,75

3) Suponga que alguien le propone un juego que consiste en lanzar al aire una moneda 3 veces, si consigue 3 caras ganará 20$, si consigue 2 caras ganará 10$, si consigue 1 cara perderá 12$ y si son todos sellos perderá 20$.
a. Determine la función de probabilidad siendo la variable aleatoria el número de caras.
b. Con base a las probabilidades obtenidas anteriormente y siendo ahora la variable aleatoria la cantidad de dinero que gana o pierde en cualquier lanzamiento. ¿Aceptaría usted jugar?

Sol: no

4) Supóngase un juego con un dado, en este juego, el jugador gana 20$ si obtiene un 2, 40$ si obtiene un 4, pierde 30$ si obtiene un 6, en tanto que ni pierde ni gana si obtiene otro resultado. Hallar la suma esperada de dinero ganado.
Sol: una ganancia de 5$.

5) En una empresa de inversiones trabajan 20 analistas. Todas las mañanas se le encarga a cada analista que evalúe de uno a cinco valores.
Los encargos que se hicieron esta mañana fueron:

# de valores	1	2	3	4	5
# de analistas	4	2	3	5	6

a. Determinar la función de probabilidad para la variable aleatoria del número de valores asignados a los analistas esta mañana.
b. Determinar la media y la varianza para la misma variable.

Sol: E(**X**) = 3,35 valores Var (**X**) = 2,23 valores^2

6) Se lanza una moneda 2 veces, represéntese por **X** el número de caras que pueden resultar.
a. Hallar la función de probabilidad para la variable aleatoria.
b. Construir la gráfica de probabilidad.
c. Hallar la función de distribución acumulativa y obtener su representación gráfica.

7) Suponga que se lanzan un par de dados honrados y que la variable aleatoria **X** denota la suma de los puntos.
a. Hallar la función de probabilidad de la variable aleatoria y construir la gráfica.
b. Hallar la función de distribución acumulativa y construir su gráfica.

8) La tabla siguiente muestra la función de distribución acumulativa de la variable aleatoria **X**

x	1	2	3	4
F(x)	1/8	3/8	3/4	1

Determinar:
a. La función de probabilidad.
b. P(1≤ **X** ≤ 3) Sol: 3/4
c. P(**X** ≥ 2) Sol: 7/8
d. P(**X** < 3) Sol: 3/8
e. P(**X** >1,4) Sol: 7/8

9) Un analista de mercado de una compañía que fabrica aviones de combate, tiene la creencia de que el nuevo avión de la compañía tiene 70% de posibilidades de ser escogido para sustituir por completo a los aviones de combate de la fuerza aérea. Sin embargo, existe una posibilidad entre 5 de que la fuerza aérea compre sólo el número necesario de esos aviones para sustituir la mitad de sus 50 aviones de combate.
Por último, existe una posibilidad entre 10 de que la fuerza aérea sustituya toda su flotilla de aviones de combate por los aviones de esta compañía y que además compre el número suficiente de éstos para ampliar el número de sus unidades en 10%.
Hallar la función de probabilidad de la variable aleatoria y trace la gráfica.

10) El director de una compañía de encomiendas, está preocupado respecto al número de cartas de 1ra clase que su compañía ha perdido. Debido a que estas cartas son transportadas en camión y por avión, el director de la compañía ha clasificado las cartas extraviadas durante los últimos 2 años como las que se perdieron en camión y las que se extraviaron en avión.
Los datos son los siguientes:

Mes / Medio	E	F	M	A	M	J	J	A	S	O	N	D
Camión	4	5	2	3	2	1	3	5	4	7	0	1
Avión	5	6	0	2	1	3	4	2	4	7	4	0

El director planea investigar a uno de los dos departamentos, el de tierra o el de aire, pero no a ambos. Si decide abrir una investigación en el departamento que tenga el mayor número esperado de cartas perdidas por mes. ¿A cuál departamento deberá investigar?
Sol: al aéreo.

11) De un estudio que se realizó en una determinada comunidad del área metropolitana de Caracas, se encontró que de cada 400 personas, 300 están alfabetizadas mientras que el resto no lo está. Si se extraen 3 personas sin reemplazo, considerando **X** como la variable que denote el total de personas alfabetizadas, construya la función de probabilidad.

12) Una moneda está cargada de tal modo que hay 3 veces más probabilidad de que caiga cara que sello. Para 3 lanzamientos independientes de la moneda, Halle:

a. La función de probabilidad de **X** siendo ésta el número total de caras.
b. Probabilidad de que cuando mucho caigan dos caras.

13) Un vendedor ambulante cuyo puesto está ubicado frente a un edificio de oficinas, tiene que determinar si en el día de hoy vende refrescos o helados; pues considera que la utilidad que realice en el día dependerá del clima. Adjunto se muestra la tabla de rendimiento:

Clima \ Venta de	Refresco	Helado
Frío	40	20
Cálido	55	80

El vendedor, con base a su experiencia, sabe que en esta época del año la probabilidad de que haga un clima cálido es de un 60%. Determine cuál de los dos bienes debe vender.
Sol: helados.

14) Un inversionista dispone de cierta cantidad de dinero para invertir de inmediato. Tiene 3 alternativas de inversión: A, B, C. En la siguiente tabla se representan las utilidades estimadas de cada cartera de acuerdo a las condiciones de la economía:

Evento	A	B	C
Economía en declive	500$	-2000$	-7000$
No hay cambios	1000$	2000$	-1000$
Economía en expansión	2000$	5000$	20000$

Con base a su experiencia, el inversionista asigna las siguientes probabilidades a cada una de las condiciones de la economía:

- Probabilidad de economía en declive: 30%
- Probabilidad de que no ocurran cambios: 50%
- Probabilidad de expansión económica: 20%

Determinar la mejor elección de cartera para el inversionista.
Sol: B.

15) Considere el lanzamiento de un par de dados, la variable aleatoria de interés representa la suma de los dos números.
Si a ese experimento se asocia un determinado juego en el cual las apuestas están basadas en los resultados de ese par de dados y se tiene que por cada apuesta que se haga se puede perder 1$ si

la suma es 5, 6, 7, 8. Se puede ganar 1$ si la suma es 3, 4, 9, 10, 11 y se puede ganar 2$ si el resultado es 2 o 12.

a. Determinar la esperanza matemática y la desviación típica de la variable aleatoria.
b. Con relación a la apuesta, ¿cuánto se espera ganar o perder?.

Sol: a. E(**X**) = 7 ; Desviación típica: 2,4152
b. Se espera una pérdida de 0,055$

16) Considere el lanzamiento de una moneda y después el de un par de dados, la variable aleatoria de interés representa la suma de los 2 números en los dados.
Si a ese experimento se asocia un determinado juego en el cual las apuestas están basadas en los resultados de ese par de dados, se puede ganar 2$ si sale cara y la suma es 3, 4, 9, 10, 11 o se pueden ganar 40$ si sale sello y el resultado es 2 o 12. De otra manera pierde 50 centavos.
¿Vale la pena jugar?
Sol: E(**X**) = 1,11$ sí vale la pena.

17) Supóngase que una variable aleatoria discreta tiene la siguiente función de probabilidades:

$$f(x) = \begin{cases} cx \text{ para } x = 1,2,3,4,5. \\ 0 \text{ en otro caso} \end{cases}$$

Determine el valor de "c"

Sol: $c = \dfrac{1}{15}$

18) Supóngase que se lanzan un par de dados equilibrados y sea "X" el valor absoluto de la diferencia entre los dos números que aparecen. Determine la función de probabilidad de "X".
Sol:

X	0	1	2	3	4	5
f(x)	1/6	5/18	2/9	1/6	1/9	1/18

19) Sea "X" una variable aleatoria discreta, que representa el número de clientes que llega a una tienda en una hora. Dada la siguiente información:

x	0	1	2	3	4	5	6	7	8
f (x)	0,05	0,10	0,10	0,10	0,20	0,25	0,10	0,05	0,05

Determinar E(X) y Var(X).
Sol: E(X) = 4 y Var(X) = 4,1

20) A pesar de encontrarse sometidos a control diario los artículos ofrecidos a la venta en un hipermercado, se estima que la probabilidad de que en un día sean vendidos "r" artículos defectuosos es: $\dfrac{2}{3}\left(\dfrac{1}{3}\right)^r \quad r = 0, 1, 2, 3, ...$

Determinar la probabilidad de que en un día sean vendidos:

a- Dos o más artículos defectuosos. Sol: 1/9
b- Cinco artículos defectuosos. Sol: 2/729
c- Tres o menos artículos defectuosos. Sol: 80/81

21) Sea "X" una variable aleatoria discreta, cuya función f(x), viene dada por:

x	1	2	3	4	5
f (x)	2/8	1/8	2/8	2/8	1/8

Determinar:

a- F (X)
b- E(X)
c- Var(X)

Ejercicios (Modelos de Variables Aleatorias Discretas)

1) Suponga que el 10% de las piezas que produce una máquina aleatoria sea defectuosa. Si se toma al azar una muestra de 20 piezas. Calcular:
a) Probabilidad de que en la muestra hayan dos piezas defectuosas Sol: 0,285
b) Probabilidad de que en la muestra hayan máximo 3 piezas defectuosas Sol: 0,867
c) Probabilidad de que en la muestra hayan 18 piezas defectuosas como mínimo. Sol: ≈ 0
d) Probabilidad de que en la muestra hayan entre 2 piezas y 5 piezas defectuosas Sol: 0,781
e) Probabilidad de que en la muestra hayan mínimo 3 piezas defectuosas Sol: 0,320

2) Suponga que en cierta población muy grande, el 52% de los nacimientos registrados son de varones. Si tomamos 5 registros, defina la variable que le permita calcular las siguientes probabilidades:
a) Que 2 registros correspondan a varones Sol: 0,2990
b) Menos de 3 sean varones Sol: 0,4625

3) Suponga que el 10% de los tornillos que produce una máquina son defectuosos. Halle la media y la varianza de **X** = número de tornillos defectuosos en una muestra de 100.
Sol: 10 y 9

4) Una urna contiene 4 esferas rojas y 6 negras, se extraen de la urna 4 esferas. Suponiendo que el muestreo se hace con reemplazo, calcular la probabilidad de que:
a) Haya a lo más una esfera roja en la muestra Sol: 0,4752
b) No haya ninguna esfera negra en la muestra Sol: 0,0256

5) El departamento de control de calidad de una empresa que fabrica pañuelos sabe que el 5% de su producción tiene algún tipo de defecto. Los pañuelos se empaquetan en cajas con 15 elementos. Calcule la probabilidad de que una caja contenga:
a) 2 pañuelos defectuosos Sol: 0,1348
b) Menos de 3 pañuelos defectuosos Sol: 0,9634
c) Entre 3 y 5 pañuelos defectuosos Sol: 0,0362
d) Ningún pañuelo defectuoso Sol: 0,4633
e) El número esperado de pañuelos defectuosos en una caja Sol: 0,7500

6) Suponga que una empresa produce 100 unidades de las cuales 90 son buenas y 10 son defectuosas. Se escogen 20 unidades sin reemplazo, halle la probabilidad de que resulten 5 defectuosas:
Sol: 0,0215

7) Una caja contiene 30 baterías para radio, de las cuales, 5 son defectuosas. De la caja se escogen al azar 6 baterías, halle la probabilidad de que:
a) 2 sean defectuosas Sol: 0,2130
b) Ninguna sea defectuosa Sol: 0,2982
c) Menos de 3 sean defectuosas Sol: 0,9586

8) Una caja tiene 6 esferas blancas y 4 rojas. Se realiza un experimento en el cual se selecciona una esfera aleatoriamente y se observa su color, pero no se reemplaza la esfera.

Hallar la probabilidad de que después de 5 pruebas del experimento se hayan escogido 3 esferas blancas: Sol: 0,4761

9) Una caja contiene 5 esferas rojas y 10 blancas. Si se seleccionan 8 esferas aleatoriamente y no hay reemplazamiento, determinar la probabilidad de que:
a) 4 sean rojas Sol: 0,1631
b) Todas sean blancas Sol: 0,0069
c) Al menos una sea roja. Sol: 0,9930

10) De 60 aspirantes a una universidad 40 son del oriente: Si se seleccionan 20 aspirantes aleatoriamente, hallar la probabilidad de que:
a) 10 sean del oriente Sol: 0,0373
b) No más de 2 sean de oriente Sol: $3,5 * 10^{-11}$

11) Suponga que el número de llamadas que llegan a un conmutador es de 0,5 por minuto en promedio, halle la probabilidad de que:
a) En un minuto lleguen más de 3 llamadas Sol: 0,0017
b) En un minuto no lleguen llamadas Sol: 0,607
c) En 3 minutos lleguen menos de 5 llamadas Sol: 0,9814
d) En 5 minutos más de 2 llamadas Sol: 0,456
e) ¿Cuántas llamadas se espera que lleguen al conmutador en cinco minutos? Sol: 2,5

12) El promedio de personas que llegan a la ventanilla de un banco por minuto durante las horas hábiles es una. Halle la probabilidad de que en un minuto dado:
a) No aparezcan clientes Sol: 0,3678
b) Hayan 3 o más clientes Sol: 0,0804
c) Hayan 3 o menos clientes Sol: 0,9810

13) Se ha determinado que en una autopista se da en promedio 10 animales vagabundos muertos por kilómetro. Halle la probabilidad de que en 100 metros:
a) Se encuentren 2 o más animales muertos Sol: 0,2643
b) No se encuentre ningún animal muerto Sol: 0,3678
c) Menos de 3 animales muertos Sol: 0,9196
d) ¿Cuántos animales muertos se espera encontrar en un trayecto de 500 metros? Sol: 5

14) Según la National Office Of Vital Statistics Of Vital Statistics Of The U.S Department Of Health , Education and Wilfare, el promedio de ahogados en accidentes por año es 3 de cada 100.000 personas. Hallar la probabilidad de que en una ciudad cuya población es de 200.000 personas ocurran:
a) Ningún ahogado por año Sol: 0,0024
b) 2 ahogados por año Sol: 0,0446
c) 6 ahogados por año Sol: 0,1606
d) 8 ahogados por año Sol: 0,1032
e) Entre 4 y 8 ahogados por año Sol: 0,696

15) Si la probabilidad de que un individuo sufra una reacción por una inyección de un determinado suero es 0,001; determinar la probabilidad de que de un total de 2000 individuos:

a) Exactamente 3 tengan reacción Sol: 0,180
b) Más de 2 individuos tengan reacción Sol: 0,323

16) Diez por ciento de las herramientas producidas en un proceso de fabricación determinado resultan defectuosas. Hallar la probabilidad de que en una muestra de 10 herramientas seleccionadas aleatoriamente exactamente 2 estén defectuosas, empleando:
a) La distribución binomial Sol: 0,19
b) La aproximación de Poisson a la distribución binomial Sol: 0,18

17) Cierto número de líneas aéreas dan servicio local entre las ciudades de Washington D.C. y New York, en los Estados Unidos. Debido al congestionamiento de tráfico en los aeropuertos de ambas ciudades, los vuelos locales se retrasan hasta 2 horas.
Datos recientes revelan que el 25% de los vuelos locales entre Washington y New York se retrasan se retrasan por más de 30 minutos. Suponga que 5 personas toman diferentes vuelos entre Washington y New York (los vuelos son independientes entre sí).
a) Encuentre la probabilidad de que las 5 personas hayan llegado a New York con un retraso menor a 30 minutos. Sol: 0,2373
b) Encuentre la probabilidad de que al menos 3 de las 5 personas hayan llegado a New York con un retraso menor a 30 minutos. Sol: 0,8964

18) El 20% de las ventas de automóviles nuevos en USA corresponde a los automóviles importados. Suponga que se seleccionan al azar 4 personas que han comprado un automóvil nuevo durante la semana pasada.
a) Encuentre la probabilidad de que las 4 personas hayan comprado un automóvil importado.
b) Encuentre la probabilidad de que sólo una de ellas haya comprado un automóvil importado.
c) Encuentre la probabilidad de que ninguna de ellas hayan comprado un automóvil importado
Sol: 0,0016 ; 0,4096 ; 0,4096 respectivamente

19) Los registros de mantenimiento revelan que solamente 1 de cada 100 máquinas de escribir de cierta marca requieren de una reparación mayor durante el 1^{er} año de uso. El gerente de una oficina ordenó la compra de 10 máquinas de esta marca.
a) Encuentre la probabilidad de que ninguna de las máquinas requieran una reparación mayor durante el 1^{er} año de uso. Sol: 0,90438
b) Encuentre la probabilidad de que 2 máquinas requieran una reparación mayor durante el 1^{er} año de uso. Sol: 0,00168

20) El centro de políticas energéticas de la agencia para la protección ambiental reporta que el 75% de las viviendas de Nueva Inglaterra utilizan quemadores de petróleo para calefacción.
Si se sabe que una comunidad de Nueva Inglaterra tiene 2500 viviendas, encuentre el número esperado de las que usan quemadores de petróleo para calefacción y la desviación estándar.
Sol: $E(\mathbf{X}) = 1875$ viviendas, $\sigma(\mathbf{X}) = 21{,}651$ viviendas

21) El reporte anual es uno de los documentos más importantes producidos por las empresas de propiedad pública y su producción representa un gasto de importancia considerable. Sin embargo, un estudio reciente revela que 40% de los accionistas dedican 5 minutos o menos a la lectura del reporte anual de su compañía. Suponga que se eligen al azar 100 accionistas de empresas de propiedad pública.

a) Encuentre el valor esperado del número de accionistas que dedican 5 minutos o menos a la lectura del reporte anual de su compañía. Sol: 40 accionistas
b) Determine la desviación estándar. Sol: 4,898 accionistas
c) Si se observa que 25 de los 100 accionistas seleccionados dedican no más de 5 minutos a la lectura del reporte anual. ¿Sería razonable pensar que la proporción de accionistas que dedican 5 minutos o menos a la lectura del reporte anual es 40%? Sol: No.

22) Suponga que una compañía de seguros vendió pólizas de seguros de vida a 5000 hombres de 42 años de edad. Si los estudios actuariales indican que la probabilidad de que un hombre de 42 años muera en un determinado año es 0,001. ¿Cuál es la probabilidad de que la compañía pague 4 indemnizaciones en un determinado año?.
Sol: 0,1755200 (Por binomial) 0,1745 (Por aproximación de Poisson)

23) La calidad aparente de un producto es por lo menos tan importante como la calidad real que se determina a través de laboratorios imparciales de prueba.
Los resultados de una investigación revelan que el 40% de los hombres de negocios de Japón piensan que los artículos eléctricos japoneses son de mayor calidad que los manufacturados en USA, pero únicamente el 5% de los hombres de negocios de USA tienen la misma opinión.
Es posible que esta diferencia se deba a la calidad aparente, usando la aproximación de Poisson a la binomial, estime la probabilidad de que exactamente 5 de una muestra de 100 hombres de negocios de USA prefieran los productos eléctricos japoneses.
Estime también, la probabilidad de que no más de 5 de los 100 prefieran los productos japoneses.
Sol: 0,1754 ; 0,6159

24) El gerente local de una compañía de renta de automóviles compra neumáticos en lotes de 500 para aprovechar los descuentos por compras al mayor.
El gerente sabe, por experiencias anteriores, que el 1% de los neumáticos nuevos adquiridos en un determinado almacén salen defectuosos y se deben reemplazar durante la 1^{ra} semana de uso.
Encuentre la probabilidad de que en un envío de 500 neumáticos haya solamente uno defectuoso; no más de 3 defectuosos y ninguno defectuoso.
Sol: 0,03368 ; 0,2650 ; 0,006737 respectivamente.

25) La central telefónica de un edificio de consultorios médicos puede manejar un máximo de 5 llamadas por minuto. Si la experiencia indica que se recibe un promedio de 120 llamadas por hora, encuentre la probabilidad de que en un determinado minuto la central esté sobrecargada.
Sol: 0,0168

26) Un determinado antibiótico se envía a las farmacias en cajas de 24 botellas, el farmacéutico sospecha que la cantidad de antibiótico en algunos frascos es deficiente y decide analizar el contenido de 5 frascos.
Suponga que 10 de las 24 botellas tienen cantidad deficiente de antibiótico.
a) Encuentre la probabilidad de que ninguno de los frascos analizados tenga una cantidad deficiente de antibiótico. Sol: 0,0471
b) Encuentre la probabilidad de que uno de los frascos analizados tenga una cantidad deficiente de antibiótico. Sol: 0,2355

27) Un embarque de 10 artículos contiene 2 unidades defectuosas y 8 no defectuosas.

Al revisarlo, se tomará una muestra y las unidades se inspeccionarán. Si se encuentra una unidad defectuosa, se rechazará todo el embarque.
a) Si se selecciona una muestra de 3 artículos. ¿Cuál es la probabilidad de rechazar el embarque?.
b) Si se selecciona una muestra de 4 artículos. ¿Cuál es la probabilidad de rechazar el embarque?
c) Si la gerencia estuviera de acuerdo en que hubiese una probabilidad de 0,47 de rechazar un embarque con 2 unidades defectuosas y 8 no defectuosas. ¿De qué tamaño se debe seleccionar la muestra?
Sol: 0,5333 ; 0,6666 ; 7

28) Los pasajeros de las aerolíneas llegan al azar e independientemente a la sección de documentación en un gran aeropuerto internacional. La frecuencia promedio de llegadas es de 10 pasajeros por minuto.
a) ¿Cuál es la probabilidad de que no lleguen pasajeros en un minuto?
b) ¿Cuál es la probabilidad de que lleguen 3 pasajeros en un minuto?
c) ¿Cuál es la probabilidad de no llegadas en período de 15 segundos?
d) ¿Cuál es la probabilidad de al menos una llegada en período de 15 segundos?
Sol: 0,000045 ; 0,0103 ; 0,0820 ; 0,918

29) Un agente de seguros vende pólizas a 5 individuos, todos de la misma edad. De acuerdo con las tablas actuariales, la probabilidad de que un individuo con esa edad viva 30 años más es de 3/5. Determinar la probabilidad de que dentro de 30 años vivan:
a) Los cinco individuos. Sol: 0,07776
b) Al menos tres. Sol: 0,68256
c) Sólo dos. Sol: 0,2304
d) Al menos uno. Sol: 0,98976

30) Una empresa, dedicada a la venta de un determinado tipo de artículo, sabe que el 20% de las unidades adquiridas de dicho artículo son bajo la forma de pago "al contado". Si en un período de tiempo determinado, se han adquirido cinco unidades, determinar la probabilidad de que:
a) Dos o más se hayan adquirido al contado. Sol: 0,2627
b) Dos o menos se hayan adquirido a plazos. Sol: 0,0579

31) El número medio de automóviles que llega a una estación de servicio de gasolina es de 210 por hora. Si dicha estación puede atender un máximo de diez automóviles por minuto, determinar la probabilidad de que en un minuto dado, lleguen a la estación de suministro más automóviles de los que puede atender.
Sol: 0,0009

32) Por parte de una compañía de seguros se sabe que el 0,003% de los individuos de una población fallece cada año de un determinado tipo de accidente. Determinar:
a) La probabilidad de que la compañía tenga que pagar a más de tres de los 10000 asegurados contra tal tipo de accidente en un año determinado. Sol: 0,0004
b) El número de accidentes esperados. Sol: 0,3

33) En una determinada zona geográfica se pretende introducir un nuevo producto del que es razonable esperar sea demandado por el 0,4% de los habitantes de dicha zona. Determinar la probabilidad de que consultados 1000 de éstos, dicho producto sea demandado:

a) Por tres ó más Sol: 0,7619
b) Por cinco ó menos. Sol: 0,7852

34) El promedio de un bateador de béisbol, es de 0,30. Si batea 4 veces. Determine la probabilidad de que logre:
a) Dos aciertos. Sol: 0,2646
b) Por lo menos un acierto. Sol: 0,7599

35) El equipo DPH tiene 2/5 de probabilidad de ganar cuando juega. Si juega 4 veces, hallar la probabilidad de que DPH gane:
a) Dos partidos. Sol: 216/625
b) Por lo menos un partido. Sol: 544/625
c) Más de la mitad de los partidos. Sol: 112/625

36) Un dado corriente se lanza 1620 veces. Hallar el número esperado de veces que sale el seis y la desviación estándar.
Sol: 270 y 15

37) Sea X una variable aleatoria binomialmente distribuida con $E(X) = 2$ y $Var(X) = 4/3$. Hallar la distribución de X.

38) Supóngase que el 1% de los artículos producidos por una máquina son defectuosos. Hallar la probabilidad de que 3 ó más artículos sean defectuosos en una muestra de 100.
Sol: 0,080

39) El proceso de admisión a un determinado plantel, consiste en someter al aspirante a dos pruebas, una teórica y otra práctica. El examen teórico, consta de cinco preguntas con tres posibles respuestas, de las cuales, una sola es la correcta en cada caso; para aprobar este examen es necesario responder acertadamente por lo menos cuatro preguntas.
Solamente, en caso de aprobar el examen teórico, el aspirante pasa el examen práctico; el cual consta de tres preguntas con dos posibles respuestas y sólo una es la correcta en cada caso. Para aprobar el examen práctico, es necesario responder acertadamente por lo menos dos preguntas y los aspirantes que aprueben ambos exámenes son admitidos.
Si un aspirante responde al azar todas las preguntas de ambos exámenes, ¿cuál es la probabilidad de que sea admitido al plantel?.
Sol: 0,0226

40) Una compañía de bienes raíces observa que uno de cada diez compradores potenciales de casas, prometen comprar una si regresan por segunda vez. En diez de estos casos, encuentre la probabilidad de que ninguno haga oferta.
Sol: 0,3487

41) Las investigaciones médicas señalan que el 20% de la población general sufre efectos negativos colaterales al ingerir un nuevo fármaco. Si un médico receta a cuatro pacientes dicho fármaco, cuál es la probabilidad de que:
a) Ninguno sufra efectos colaterales. Sol: 0,4096
b) Al menos uno presente tales efectos. Sol: 0,5904

c) Todos los tengan. Sol: 0,0016

42) De diez empleados de cierta empresa, siete tenían esposas que también trabajan fuera del hogar. ¿Cuál es la probabilidad de que a lo sumo un esposo tenga una esposa que esté empleada fuera de casa si se seleccionan tres trabajadores al azar?
Sol: 0,1833

43) En K.A.O. se acaba de recibir un embarque de 10 televisores. Poco después de haberse realizado la entrega, el fabricante llamó para informar que por error se habían enviado tres televisores defectuosos. El propietario de la empresa, decidió probar dos televisores de los 10 recibidos. ¿Cuál es la probabilidad de que ninguno de los dos tenga defectos?.
Sol: 0,4667

44) Una distribuidora de bebidas tiene 15 camiones de reparto. Supóngase que 6 de los mismos tienen problemas con los frenos, si se seleccionan al azar cinco camiones para probarlos, ¿cuál es la probabilidad de que dos de los vehículos examinados tengan problemas con los frenos.
Sol: 0,4196

45) Se estima que el 0,5% de las llamadas al departamento de facturación de CANTV reciben la señal de ocupado. ¿Cuál es la probabilidad de que de las 1200 llamadas del día de ayer, por lo menos cinco hayan recibido dicha señal?.
Sol: 0,7149

46) Si un banco recibe en promedio 6 cheques sin fondo por día. Determine las probabilidades de que reciba:
a) Cuatro cheques sin fondo en un día dado. Sol: 0,13392
b) Diez cheques sin fondo en cualquiera de dos días consecutivos. Sol: 0,1

47) En la inspección de hojalata producida por un proceso electrolítico continuo, se identifican 0,2 imperfecciones en promedio por minuto. Determine las probabilidades de identificar:
a) Una imperfección en 3 minutos. Sol: 0,329307
b) Al menos dos imperfecciones en 5 minutos. Sol: 0,26416
c) Cuando más una imperfección en 15 minutos. Sol: 0,19921

48) El número promedio de ratas de campo por acre, en un campo de trigo de 5 acres, se estima que es de 12. Encuentre la probabilidad de que menos de 7 ratas de campo se encuentren
a) En un acre de terreno determinado. Sol: 0,0458
b) En 2 de los siguientes 3 acres inspeccionados. Sol: 0,0060

49) La probabilidad de que una persona que vive en una cierta ciudad posea un perro se estima en 0,3. Encuentre la probabilidad de que la décima persona entrevistada aleatoriamente en esta ciudad sea la quinta persona que posee un perro.
Sol: 0,0515

50) Un científico inocula varios ratones, uno a la vez, con un germen de una enfermedad hasta que obtiene 2 que la han contraído. Si la probabilidad de contraer la enfermedad es $1/6$, ¿cuál es la probabilidad de que se requieran 8 ratones?
Sol: 0,0651

51) Suponga que la probabilidad de que una persona determinada crea una historia acerca de los atentados a una famosa actriz es 0,8. ¿Cuál es la probabilidad de que:
(a) La sexta persona que escucha tal historia sea la cuarta que la crea? Sol: 0,1638
(b) La tercera persona que escucha tal historia sea la primera en creerla? Sol: 0,032

52) Encuentre la probabilidad de que una persona que lanza una moneda obtenga:
(a) La tercera cara en el séptimo lanzamiento. Sol: 0,1172
(b) La primera cara en el cuarto lanzamiento. Sol: $1/16$

53) Tres personas lanzan una moneda y la que salga dispareja paga los cafés. Si todas las monedas caen iguales, se lanzan nuevamente. Encuentre la probabilidad de que se necesiten menos de 4 lanzamientos.
Sol: 0,9843

54) De acuerdo con un estudio publicado por un grupo de sociólogos de la Universidad de Massachussets, alrededor de las dos terceras partes de los 20 millones de personas en USA que consumen valium son mujeres. Suponiendo que ésta es una estimación válida, encuentre la probabilidad de que en un determinado día la quinta receta médica por valium sea:
(a) La primera prescripción de valium para una mujer. Sol: $2/243$
(b) La tercera prescripción de valium para una mujer. Sol: $16/81$

55) La probabilidad de que un estudiante para piloto apruebe el examen escrito para obtener su licencia para piloto privado es 0,7. Encuentre la probabilidad de que una persona apruebe el examen:
(a) En el tercer intento. Sol: 0,0630
(b) Antes del cuarto intento. Sol: 0,9730