**CAPÍTULO 1**

**FUNDAMENTOS TEÓRICOS**

**1.1. INTRODUCCIÓN**

“Un informe técnico es la exposición por escrito de las circunstancias observadas en el reconocimiento de precios, edificaciones, documentos, etc., o en el examen del asunto que se considera, con explicaciones técnicas, económicas, etc” **(1)**.

Dicho en otras palabras, es un documento que describe el progreso o resultados de una investigación científica o técnica, o el estado de un problema científico.

**1.2. DEFINICIONES BÁSICAS DE ESTADÍSTICA**

* **Parámetro:** Es una medida de resumen que se calcula para describir una característica de toda una población.
* **Población:** Es la totalidad de elementos o cosas bajo consideración.
* **Muestra:** Es la porción de la población que se selecciona para su [análisis](http://www.monografias.com/trabajos11/metods/metods.shtml#ANALIT).
* **Frecuencia:** Es el número de veces que se repite un [valor](http://www.monografias.com/trabajos14/nuevmicro/nuevmicro.shtml), dato o término dentro de una serie en estudio**.**
* **Variable:** Es la [caracter](http://www.monografias.com/trabajos10/carso/carso.shtml)ística de [interés](http://www.monografias.com/trabajos7/tain/tain.shtml) sobre cada elemento de una [población](http://www.monografias.com/trabajos/explodemo/explodemo.shtml) o [muestra](http://www.monografias.com/trabajos11/tebas/tebas.shtml) y puede tomar diferentes [valores](http://www.monografias.com/trabajos14/nuevmicro/nuevmicro.shtml).
* **Datos:** Son números o medidas que han sido recopilados como resultado de observaciones.
* **Tablas estadísticas:** Son aquellas que están formadas por la columna [matriz](http://www.monografias.com/trabajos10/macroecon/macroecon.shtml) y el cuerpo esta compuesto por más de una columna y se dividen en simples y complejas.
* **Gráficos estadísticos:** Son datos cuantitativos que vienen representados por [dibujos](http://www.monografias.com/trabajos13/histarte/histarte.shtml#ORIGEN) geométricos donde la longitud o el área de una parte de la figura es proporcional a la cantidad o magnitud representada.
* **Escala:** Es la asociación de cosas distintas pero de la misma especie. Es el tamaño o proporción con el que se desarrolla un [plan](http://www.monografias.com/trabajos7/plane/plane.shtml) de ideas.

**1.3. PRESENTACIÓN DE DATOS NUMÉRICOS EN TABLAS Y DIAGRAMAS**

**1.3.1. Distribución de frecuencia**

“Es una tabla de resumen en la que los datos se disponen en agrupamientos o categorías convenientemente establecidas de clases ordenadas numéricamente.

En esta forma las características más importantes de los datos se aproximan muy fácilmente, compensando así el hecho de que cuando los datos se agrupan de ese modo, la información inicial referente a las observaciones individuales de que antes se disponía se pierde a través del proceso de agrupamiento o condensación.

Al construir la tabla de frecuencia-distribución, debe ponerse atención a:

1. Seleccionar el número apropiado de agrupamientos de clase para la tabla.
2. Obtener un intervalo o ancho de clase de cada agrupamiento de clase.
3. Establecer los límites de cada agrupamiento de clase para evitar los traslapes.

La principal ventaja de usar una de estas tablas de resumen es que las principales características de los datos se hacen evidentes inmediatamente para el lector.

La principal desventaja de tal tabla de resumen es que no podemos saber como se distribuyen [los valores](http://www.monografias.com/trabajos14/nuevmicro/nuevmicro.shtml) individuales dentro de un intervalo de clase particular sin tener acceso a los datos originales. El punto medio de la clase, sin embargo, es el valor usado para representar todos los datos resumidos en un intervalo particular.

El punto medio de una clase (o marca de clase) es el punto a la mitad de los límites de cada clase y es representativo de los datos de esa clase.

Tabulación de datos numéricos: La distribución de frecuencia relativa y distribución de porcentaje, donde la distribución relativa de frecuencia se forma dividiendo las frecuencias de cada clase de distribución de frecuencia entre el número total de observaciones. Entonces puede formarse una distribución de porcentaje multiplicando cada frecuencia relativa o proporción entre 100.

La distribución de frecuencia relativa o la distribución de porcentaje se vuelve esencial siempre que una serie de datos se compara con otra seria de datos, especialmente si difiere el número de observaciones en cada serie de datos” **(2)**.

**1.3.2. Graficación de datos numéricos: el histograma y el polígono**

**1.3.2.1. Histogramas**

“Los histogramas son diagramas de barras verticales en los que se construyen barras rectangulares en los límites de cada clase. La variable aleatoria o fenómeno de interés se despliega a lo largo del eje horizontal; el eje vertical representa el número, proporción o porcentaje de observaciones por intervalo de clase, dependiendo de si el histograma particular, es un histograma de frecuencia, un histograma de frecuencia relativa o histograma de porcentaje

Al comparar dos o más series de datos, ni los diagramas de tallo y hoja ni los histogramas pueden construirse en la misma gráfica. Con respecto a estos últimos, la sobreposición de barras verticales de uno en el otro ocasionaría dificultades de interpretación; en estos casos se usan los polígonos”. **(2)**

**1.3.2.2. Polígonos**

“El polígono de porcentaje se forma permitiendo que el punto medio de cada clase represente los datos de esa clase y luego conectando la sucesión de puntos [medios](http://www.monografias.com/trabajos14/medios-comunicacion/medios-comunicacion.shtml) con sus respectivos porcentajes de clase”. **(2)**

**1.3.2.3. Distribuciones acumulativas y polígonos acumulativos**

“Una tabla de distribución de porcentaje acumulativo se construye registrando primero los límites inferiores de cada clase a partir de la distribución de porcentaje y luego insertando un límite extra al final”. **(2)**

**1.3.2.4. Polígono de porcentaje acumulativo**

“Para construir un polígono de porcentaje acumulativo (también llamado ojiva), el fenómeno se grafica en el eje horizontal, mientras que los porcentajes acumulativos se grafican en el eje vertical”. **(2)**

**1.4. PROPIEDADES DE LOS DATOS NUMÉRICOS.**

“Las tres mejores propiedades que describe una serie numérica de datos son:

1. Tendencia central
2. Variación
3. Forma

Si estas mediciones se calculan a partir de una muestra, se denominan estadísticas, si se calculan a partir de los datos de una población se denominan parámetros”. **(2)**

**1.4.1. Mediciones de tendencia Central**

**1.4.1.1. Media Aritmética**

“La media aritmética, es el promedio. Se calcula sumando todas las observaciones y luego dividiendo el total entre el número de elementos involucrados.

n

 **Σ** Xi

 i=1
 X =

 n

La media actúa como punto de equilibrio de tal forma que las observaciones menores compensan a las observaciones que son mayores.

La media aritmética se ve afectada en gran medida por valores extremos”. **(2)**

**1.4.1.2. La Mediana**

“La mediana es el valor central de la variable, es decir, supuesta la muestra ordenada en orden creciente o decreciente, el valor que divide en dos partes la muestra.

Para calcular la mediana debemos tener en cuenta si la variable es discreta o continua.

Cálculo de la mediana en el caso discreto:

Tendremos en cuenta el tamaño de la muestra.

Si N es Impar, hay un término central, el término que será el valor de la mediana.

Si N es Par, hay dos términos centrales, la mediana será la media de esos dos valores”. **(3)**

**1.4.1.3. La Moda**

“Es el valor de una serie de datos que aparece con más frecuencia. La moda no se ve afectada por la ocurrencia de cualquier valor extremo”. **(2)**

**1.4.1.4. Cuartiles**

“Los cuartiles sonmediciones descriptivas que dividen los datos ordenados en cuatro cuartos.

 a) Primer cuartil (Q1): Aquel valor de una serie que supera al 25% de los datos y es superado por el 75% restante.

b) Segundo cuartil (Q2): Coincide, es idéntico o similar al valor de la Mediana (Q2 = Md). Es decir, supera y es superado por el 50% de los valores de una Serie.

c) Tercer cuartil (Q3): Aquel valor, termino o dato que supera al 75% y es superado por el 25% de los datos restantes de la Serie”. **(2)**

**1.4.2. Mediciones de la Variación**

“La variación es la cantidad de dispersión o propagación en los datos”. **(2)**

**1.4.2.1. El rango**

“Es la diferencia entre la mayor y la menor observación en una serie de datos. El rango mide la propagación total en la serie de datos. La debilidad del rango es que no logra tomar en cuenta la forma en que los datos se distribuyen realmente entre el mayor y el menor valor. Sería impropio usar el rango como una medición cuando uno de o ambos componentes son observaciones extremas”. **(2)**

**Rango = Xmayor -Xmenor**

**1.4.2.2. El rango intercuartil**

“Es la diferencia entre el tercer y primer cuartil. No se ve influida por valores extremos”. **(2)**

**Rango Intercuartil = Q3 -Q1**

**1.4.2.3. La varianza y la desviación estándar**

“A diferencia de las mediciones anteriores la varianza y la desviación estándar toman en cuenta como se distribuyen las observaciones. La Varianza de muestra es el promedio de las diferencias cuadradas entre cada una de las observaciones de una serie de datos y la media. La desviación estándar es simplemente la raíz cuadrada de la varianza. La varianza y la desviación miden la dispersión promedio alrededor de la media, es decir, como las observaciones mayores fluctúan por encima de ésta y como las observaciones menores se distribuyen por debajo de ésta”. **(2)**

**1.4.2.4. El Coeficiente de Variación**

“Es una medida relativa de variación. Se expresa como porcentaje antes que en términos de las unidades de los datos particulares. Mide la dispersión en los datos relativa a la media.

El coeficiente de variación es útil al comparar la variabilidad de dos o más series de datos que se expresan en distintas unidades de medición”. **(2)**

**1.4.3. Forma**

“Para describir la forma sólo se necesita comparar la media y la mediana. Si estas dos mediciones son iguales, por lo general se puede considerar que los datos son simétricos. Si la media excede a la mediana, los datos pueden describirse de sesgo positivo o sesgadas a la derecha. Si la media es excedida por la mediana, estos datos pueden llamarse de sesgo negativo o sesgadas a la izquierda. El sesgo positivo surge cuando la media se incrementa en algunos valores inusualmente altos, el sesgo negativo ocurre cuando la media se reduce en algunos valores extremadamente bajos”. **(2)**

**1.5. PRESENTACIÓN DE DATOS CATEGÓRICOS EN TABLAS Y DIAGRAMAS**

**1.5.1. Graficación de datos categóricos de barras, de pastel y de punto**

“En la gráfica de barras, cada categoría se describe mediante una barra, cuya longitud representa la frecuencia o porcentaje de observaciones que caen en una categoría. Para construir una gráfica de barras se hacen las siguientes sugerencias:

1. Las barras deben construirse horizontalmente.
2. Todas las barras deben tener el mismo ancho.
3. Los espacios entre las barras deben variar entre la mitad del ancho de una barra hasta el ancho de una barra.
4. Las escalas y guías son auxiliares útiles en la lectura de una gráfica y deben incluirse. El punto cero u origen debe indicarse.
5. Los ejes deben etiquetarse”. **(2)**

**1.5.2. Graficación de datos categóricos: el Diagrama de Pareto.**

“El diagrama de Pareto es un tipo especial de gráfica de barras verticales en la que las respuestas categorizadas se grafican en el orden de rango descendiente de sus frecuencias y se combinan con un polígono acumulativo en la misma escala. El principio básico detrás de este dispositivo gráfico es su capacidad de distinguir los "pocos vitales" de los "muchos triviales". **(2)**

**1.5.3. Tabulación de datos categóricos**

**1.5.3.1. Tabla de contingencias y supertablas.**

“Las tablas de contingencia se usan para examinar las respuestas a dos variables categóricas simultáneamente.

Una supertabla es esencialmente una colección de tablas de contingencia, cada una con las mismas variables y categorías de columna. Sin embargo, se incluyen tantas variables de fila como se deseen para comparaciones frente a la variable de columna”. **(2)**