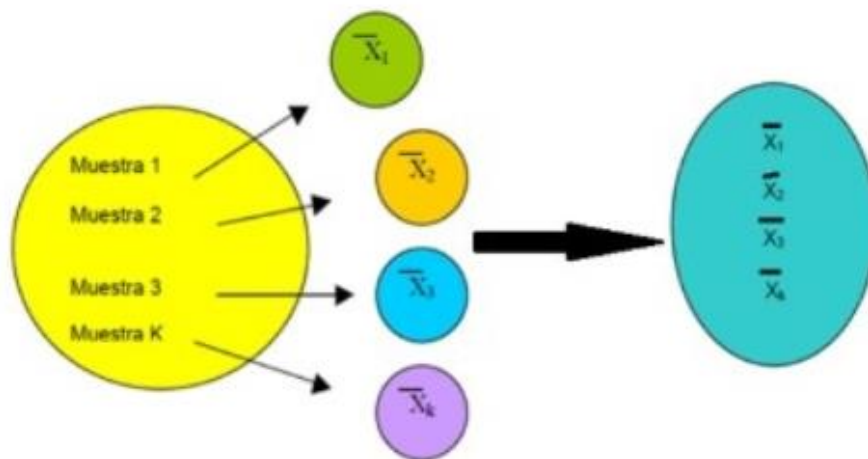


Análisis de graficas:

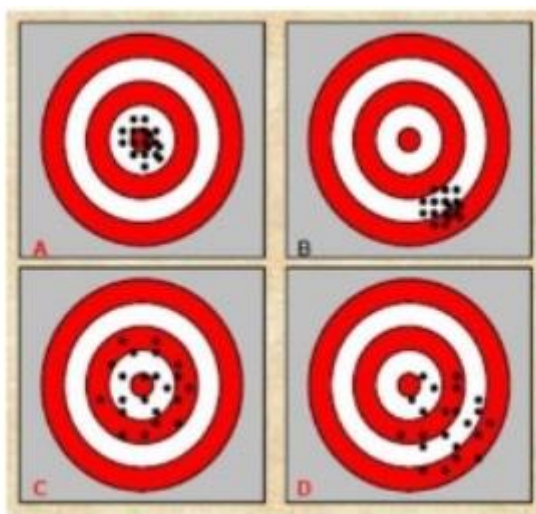
1) Tenemos una población de referencia que nos interesa estudiar y se realizan diferentes muestreos de la misma. Hay que tener en cuenta que todas las muestras van a ser distintas, del mismo tamaño y de la misma población de estudio. De cada una de esas muestras puedo tener todos los estadísticos, que en este caso particular tenemos la media, y esa distribución será la distribución muestral de la media. Todo esto se toma como una variable aleatoria.



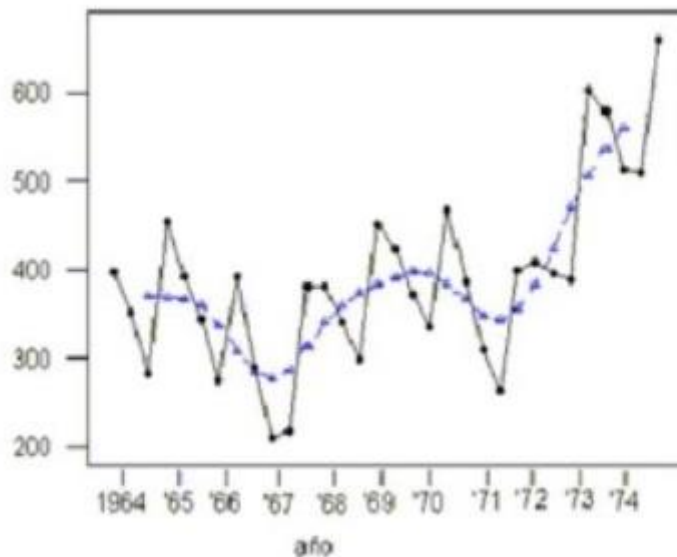
2) Lo que se puede interpretar en la siguiente figura es un estimador puntual que es una función de valores observados (muestra) que no depende de ningún parámetro desconocido.

Puede ser Insesgado, eficiente y centrado.

- A) Estimador centrado y eficiente.
- B) Estimador sesgado y eficiente.
- C) Estimador centrado e Ineficiente
- D) Estimador Sesgado e Ineficiente

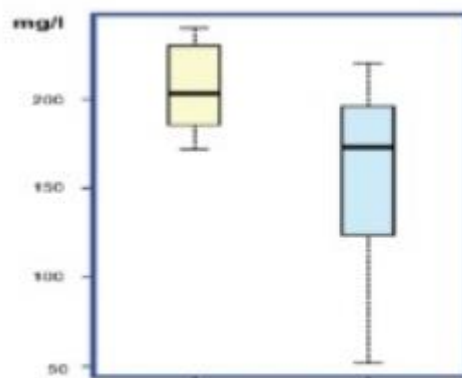


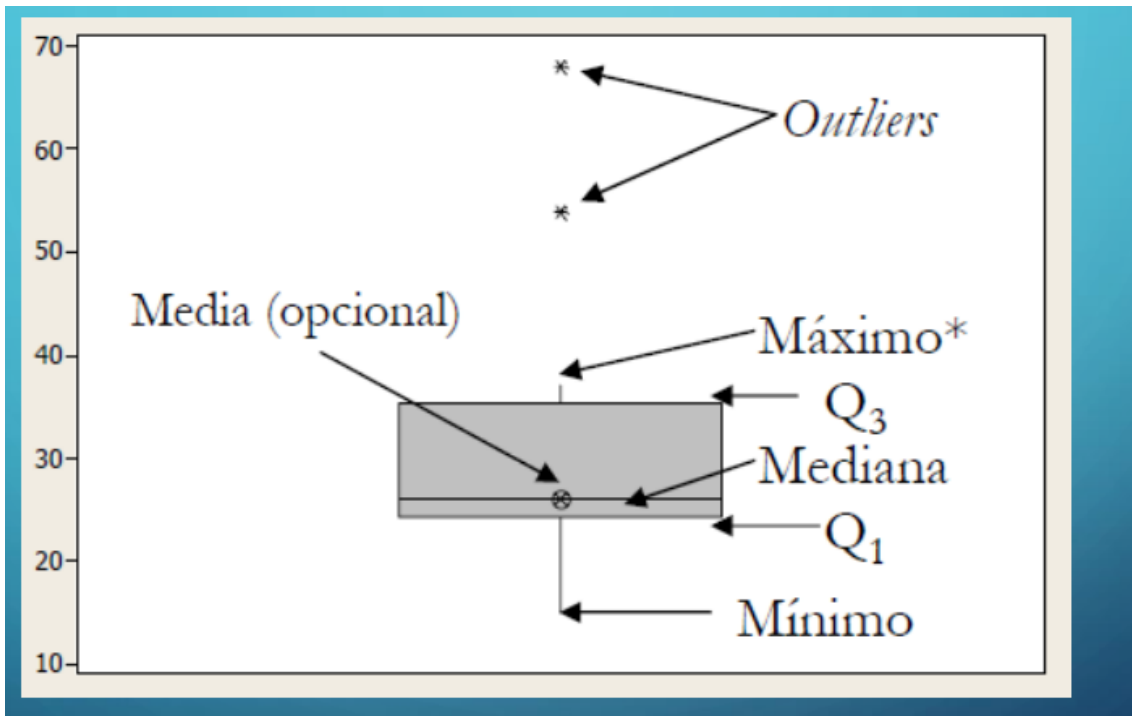
3) La figura a continuación representa una la grafica de una serie, y la misma serie suavizada. Una serie son aquellas que están formadas por valores de una variable observada a intervalos regulares de tiempo. Y suavizada quiere decir que a partir de la serie observada se forma una nueva serie en la que suavizan los efectos ajenos a la tendencia (estacionalidad, efectos aleatorios), de manera que se pueda determinar claramente la tendencia.



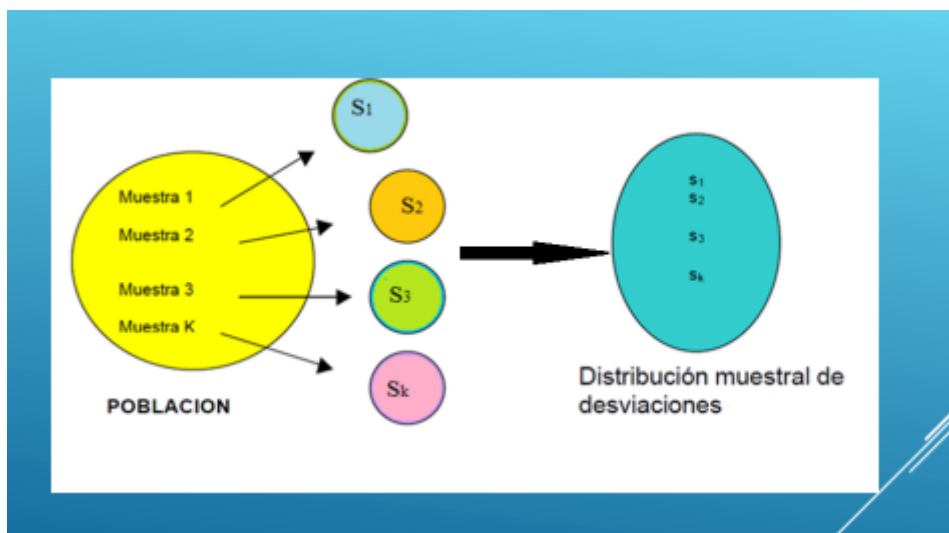
4) Lo que estamos viendo a continuación es un diagrama de tallo y hoja, y otro diagrama de caja y bigote, o boxplot. El diagrama de caja y bigote podemos ver que la línea del medio de la caja es la mediana, la “base” de la caja es el Cuartil1, y el “techo” de la caja representa el Cuartil3. Caja cortita, varianza chica. Si la mediana esta en el medio los valores son simétricos, si la mediana esta abajo los valores se acercan mas al cuartil1 si es más arriba al cuartil3

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





5) La varianza muestral es muy importante. Es tratada como VA, por lo que hay que estudiar su esperanza, su varianza y su distribución de prob. Sacamos muestras del mismo tamaño de una misma población, y en cada una de ellas calculamos el S^2 y obviamente S (desvió), y eso va a formar la distribución muestral de desviaciones.



- Notaciones en **MUESTRA** y **POBLACIÓN**

	Estadístico	Parámetro
Media	\bar{x}	μ
Varianza	s^2	σ^2
Desviación estándar	s	σ
Covarianza	s_{xy}	σ_{xy}
Correlación	r_{xy}	ρ_{xy}

HISTOGRAMA

Es una manera sencilla de representar una gran masa de datos y debe ser el comienzo de cualquier estudio más sofisticado y en el que pueden observarse tres propiedades esenciales de una distribución: forma, tendencia central o acumulación y dispersión o variabilidad. Este se obtiene por graficar en el eje x las clases y en el eje y las frecuencias. La altura de las barras del histograma tienen distinta significación según el ancho de clase sea constante o no. En el primer caso se representan frecuencias, o sea la cantidad de valores en cada clase; en el segundo caso densidad de frecuencias, o sea es el promedio, en cada clase, de cuántos valores hay por unidad de ancho de clase: $f_i/c_i = h$. En este caso el área de cada rectángulo es proporcional a la frecuencia. A diferencia del diagrama de barras, los rectángulos verticales, se representan contiguos para reflejar la idea de que la variable es continua. La forma del histograma refleja propiedades importantes de la variable estadística a la que se refiere. El número de clases o intervalos y la longitud que se consideran, depende de cada problema y de la utilización que se quiera dar a las tablas estadísticas. Lo normal es que todos los intervalos sean de la misma amplitud ($L_i - L_{i-1}$), aunque pueden existir múltiples razones donde se aconseje tomar intervalos de amplitud variable, como puede ser el caso en el que existan uno o dos intervalos donde se concentren la mayoría de los datos. La construcción de los intervalos de clase, introduce algunas cuestiones subjetivas, como son: 1) ¿Cuántos intervalos construir? Aunque no existe una regla general para usar, es evidente que el número de intervalos debe ser mayor al aumentar el tamaño muestral, lo ideal entre 5 y 20. 2) ¿Qué valor se elige como extremo inferior del primer intervalo L_0 ? Se toma como L_0 un valor "un poco menor" que el mínimo de la muestra (o el mínimo). Es muy importante hacer una buena elección de la cantidad de clases a utilizar. Para este fin se utilizan distintas reglas, una de ellas consiste en tomar el número de clases igual al entero más próximo a la raíz cuadrada del número de observaciones que se estudian, N y no ser inferior a 5 ni superior a 20, ya que en el primer caso se produciría una concentración de datos que no sería representativa de la muestra, y en el segundo caso podrían quedar intervalos vacíos, en los cuales no habría ningún valor.

Consejos: 1. Usar intervalos de la misma longitud 2. Los intervalos no pueden solaparse 3. Cada observación sólo puede pertenecer a un intervalo 4. Todos los datos deben pertenecer a algún intervalo 5. La forma del histograma depende de la amplitud del intervalo que se elija.

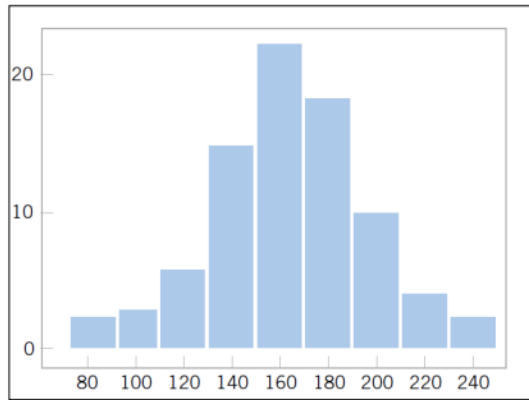
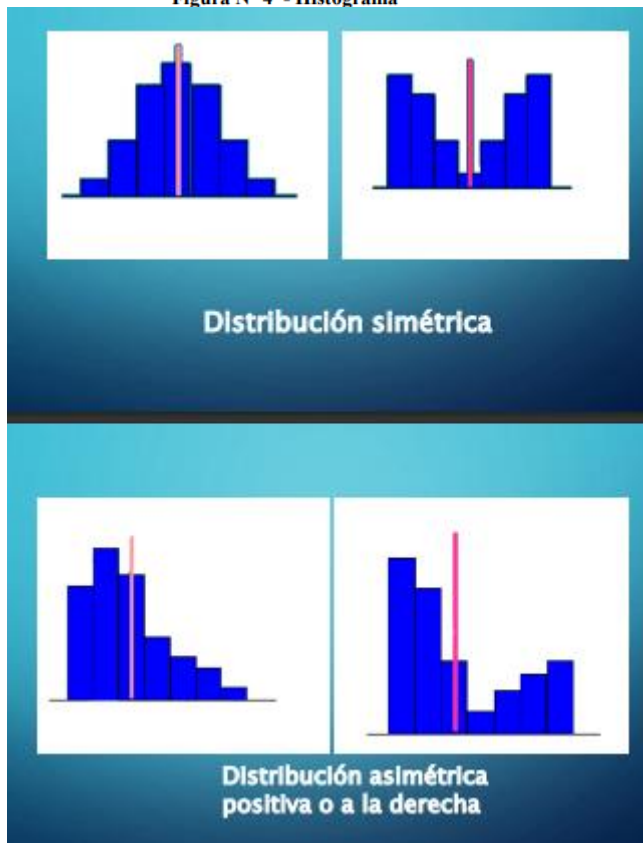


Figura N° 4 - Histograma



POLÍGONO DE FRECUENCIAS

Se construye fácilmente una vez representado el histograma, y consiste en unir los puntos del histograma que corresponden a las marcas de clase de cada intervalo mediante una recta. El diagrama, para variables continuas, se denomina polígono de frecuencias acumulado u ojiva. En estos polígonos obtenidos se aprecian con claridad propiedades importantes, da idea aproximada de qué curva teórica le correspondería a la población de la cual se obtuvo la muestra. Si las frecuencias se expresan como proporciones - es decir, divididas por el total de observaciones en la muestra lo que se obtiene es una distribución de frecuencias relativas. Cuando se realice el histograma en este caso, el área total de las barras será igual a 1

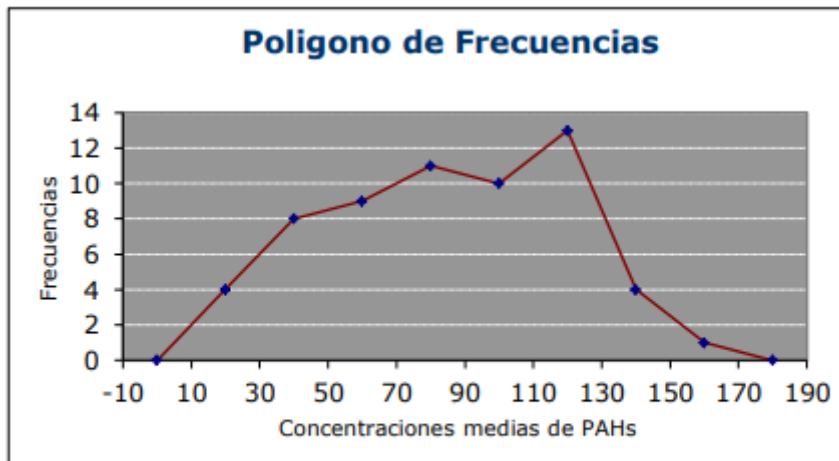
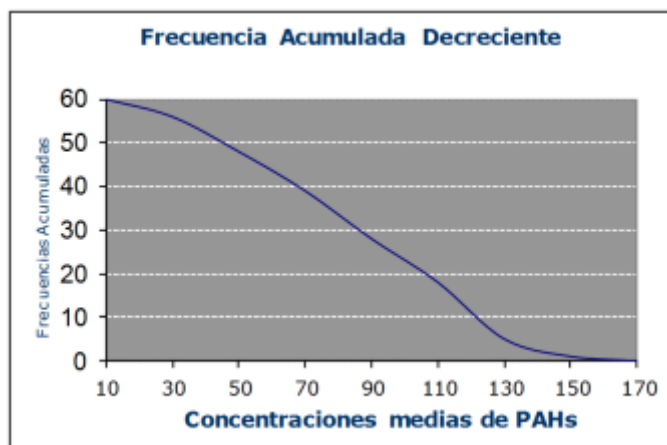
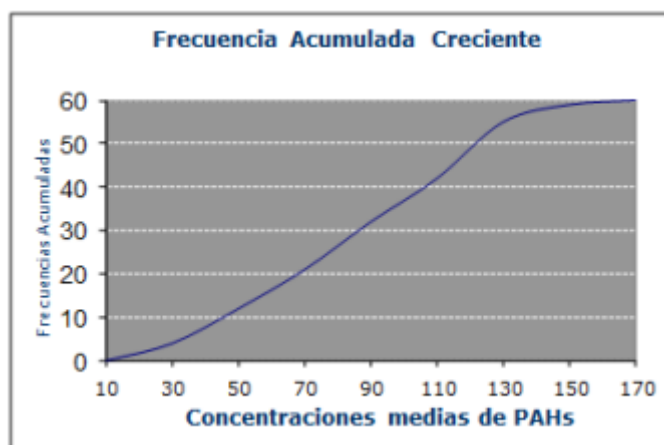
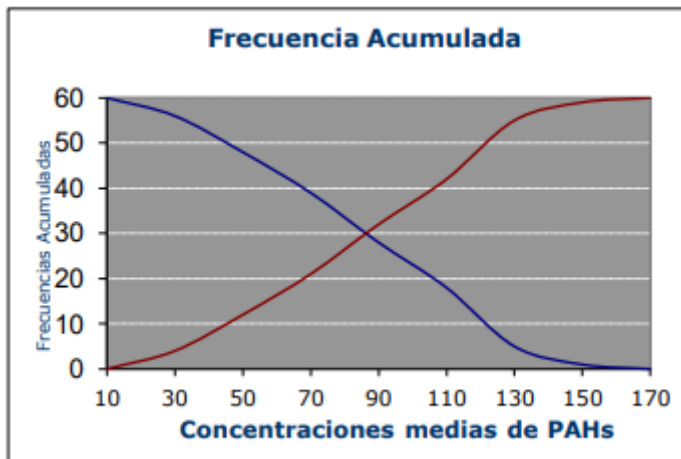


Figura N° 5 – Polígono de Frecuencias





Polígono de frecuencias acumuladas: se utilizan en variables continuas. El eje de abscisas se construye igual que en los histogramas, pero en el de ordenadas se incluyen las frecuencias acumuladas, ya sean absolutas o relativas. Sobre cada límite se levanta una perpendicular cuya longitud sea idéntica a la frecuencia acumulada y se unen los extremos superiores de dichas perpendiculares.

TALLO Y HOJA

El objetivo del diagrama de tallo y hojas es mostrar la frecuencia con la que ocurren los valores dentro de un conjunto de datos lo cual es muy parecido a lo que hace un histograma pero la diferencia es que en el diagrama de tallo y hojas no se observan barras sólidas sino que son los mismos números los que dan forma al diagrama. Puede definirse como un híbrido que combina los aspectos visuales del histograma con la información numérica que proporciona una tabla de distribución de frecuencias.

Este diagrama se construye colocando en una columna todos los números que conforman los datos eliminando la última cifra, es decir las unidades. Esta columna debe ordenarse de menor a mayor. A la derecha de cada número se escribe la última cifra o unidad de cada dato que comienza con ese número. Luego se ordenan de menor a mayor los números de cada fila. Cada valor se subdivide en tres componentes el más significativo, o sea el situado más a la derecha, se usa para formar el tallo, el segundo en significación forma la hoja, que servirá para generar un histograma y con ello proporcionar una idea de la forma de la variable, y el tercero, si existe, que es el menos significativo, se puede despreciar. Cuando existen valores muy separados del conjunto, se puede simplificar el gráfico eliminando las filas sin hojas e indicando los valores altos o bajos completos precedidos de esas palabras, ALTOS, si son muy elevados, o BAJOS si se da la circunstancia contraria. La elaboración de un gráfico de tallo y hojas es muy sencilla, y se puede considerar como la técnica de representación gráfica recomendable para variables cuantitativas, por encima de otra forma muy usual como el histograma. Se construye de la siguiente manera: 1. Ordenar el lote de datos en magnitud creciente. 2. Seleccionar un par conveniente de dígitos que permita fraccionar en dos partes el lote de datos según la característica de los datos o lo que se quiere mostrar. 3. Formar el tallo y las hojas con las fracciones respectivas. 4. Construir el tallo escribiendo verticalmente los dígitos enteros entre el 22 y 31, asociando a cada uno su hoja respectiva. Los dígitos del tallo están separados de los dígitos de la hoja por medio de una línea vertical. En términos generales un diagrama de esta naturaleza hace visibles las siguientes características: 1. Muestra el rango de valores que los datos cubren. 2. Determina donde se concentran la mayoría de los datos 3. Describen la simetría del conjunto de datos. 4. Identifica si existen huecos en la distribución de los datos. 5. Señala aquellos valores que claramente se desvían del conjunto de datos. Otra opción que presenta el

diagrama de tallo y hoja es la comparación entre dos lotes de datos, aspecto que no considera el histograma. A esta derivación se le llama diagrama de tallo y hoja en espejo

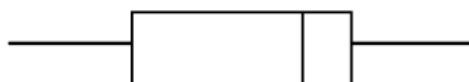
7	6	1
8	7	1
9	7	1
10	5 1	2
11	5 8 0	3
12	1 0 3	3
13	4 1 3 5 3 5	6
14	2 9 5 8 3 1 6 9	8
15	4 7 1 3 4 0 8 8 6 8 0 8	12
16	3 0 7 3 0 5 0 8 7 9	10
17	8 5 4 4 1 6 2 1 0 6	10
18	0 3 6 1 4 1 0	7
19	9 6 0 9 3 4	6
20	7 1 0 8	4
21	8	1
22	1 8 9	3
23	7	1
24	5	1

Figura N° 7 – Diagrama de Tallo y Hojas

Gráfico de Caja y Bigote - Box Plot

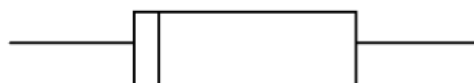
Un diagrama de caja y bigotes es una gráfica basada en cuartiles, que ayuda a visualizar un conjunto de datos. Para construir un diagrama de caja se necesita el valor mínimo, el primer cuartil, la mediana, el tercer cuartil, y el valor máximo. El cuadro encierra el rango intercuartil con el borde izquierdo (o más bajo) en el primer cuartil, Q1, y el borde derecho (o superior) en el tercer cuartil, Q3. Se traza una línea a través de la caja en el segundo cuartil (que es el percentil 50 o la mediana); una línea, o el bigote, se extiende desde cada extremo de la caja. El bigote inferior es una línea desde el primer cuartil hasta el punto de datos más pequeño dentro de 1,5 rango intercuartil. El bigote superior es una línea desde el tercer cuartil hacia los valores más grandes a 1,5 rango intercuartil. Datos que estén más lejos de estos bigotes se representan como puntos individuales. Un punto más allá de un bigote, pero menos de 3 rango intercuartil desde el borde de la caja, es llamado un caso atípico. Un punto más allá de Estadística - Ingeniería en Informática – FICH – UNL Página 14 3 rangos intercuartil del borde de la caja se llama un extremo atípico. Diferentes símbolos, tales como círculos abiertos y llenos, se utilizan en ocasiones para identificar los dos tipos de valores atípicos

Si la mediana está ubicada como sigue



entonces la distribución es asimétrica negativa. (izquierda)

Si la mediana está de esta manera



Entonces la distribución es asimétrica positiva. (derecha)

ASIMETRIA

Cuando la distribución es simétrica la media, mediana y el modo coinciden. Cuando es asimétrica esos valores difieren. Como se ha visto la media aritmética es el valor de la tendencia central más afectado por los valores extremos, es por esto que cuanto mayor sea la distancia entre la media y el modo mayor será el grado de asimetría. Esta diferencia entre media y modo se suele usar como medida de asimetría ya que cuanto mayor sea esta distancia mayor será la asimetría. A los fines de comparar la asimetría entre dos distribuciones y salvar la diferencia de unidades y la diferencia en las dispersiones es que se divide por el desvío. Debido a que el modo se encuentra aproximadamente algunas veces, se prefiere trabajar con la mediana que se encuentra mejor y teniendo en cuenta la relación vista entre la media, mediana y modo, se pueden obtener las siguientes expresiones para la asimetría:

$$As = \frac{(\bar{x} - \text{Modo})}{S}$$

$$As = \frac{3(\bar{x} - \text{Mediana})}{S}$$

Este valor sería aproximadamente igual a 0 para una distribución simétrica, positivo para una distribución asimétrica hacia la izquierda y negativo para una distribución asimétrica hacia la derecha

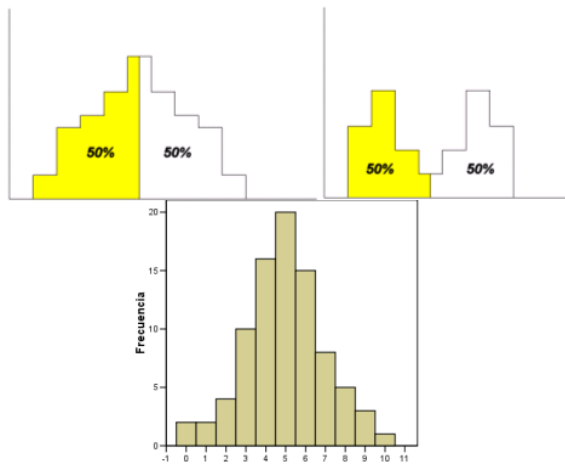


Figura N° 9 – Distribución simétrica

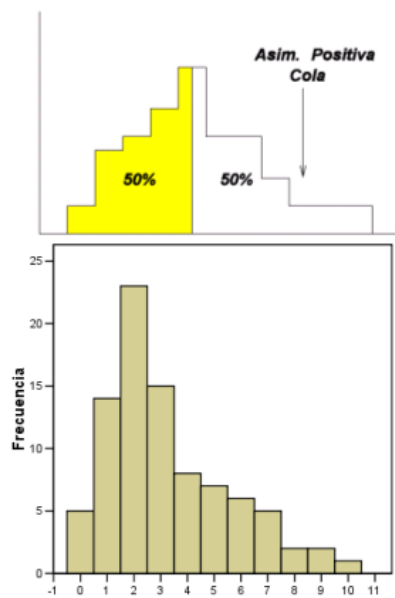
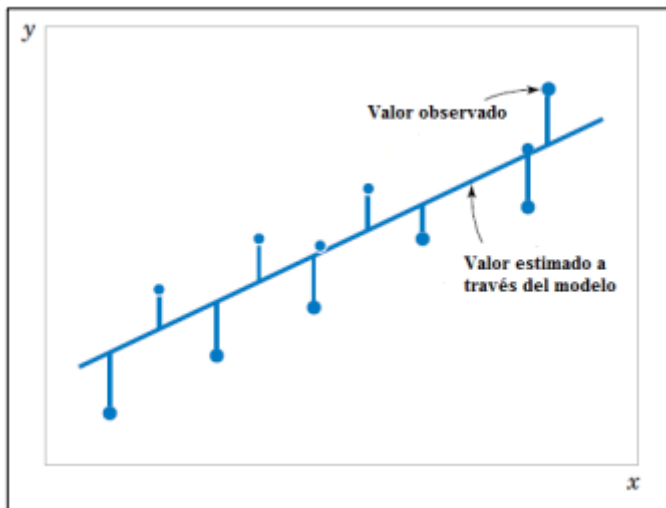


Figura N° 10 – Distribución asimétrica positiva o a la derecha

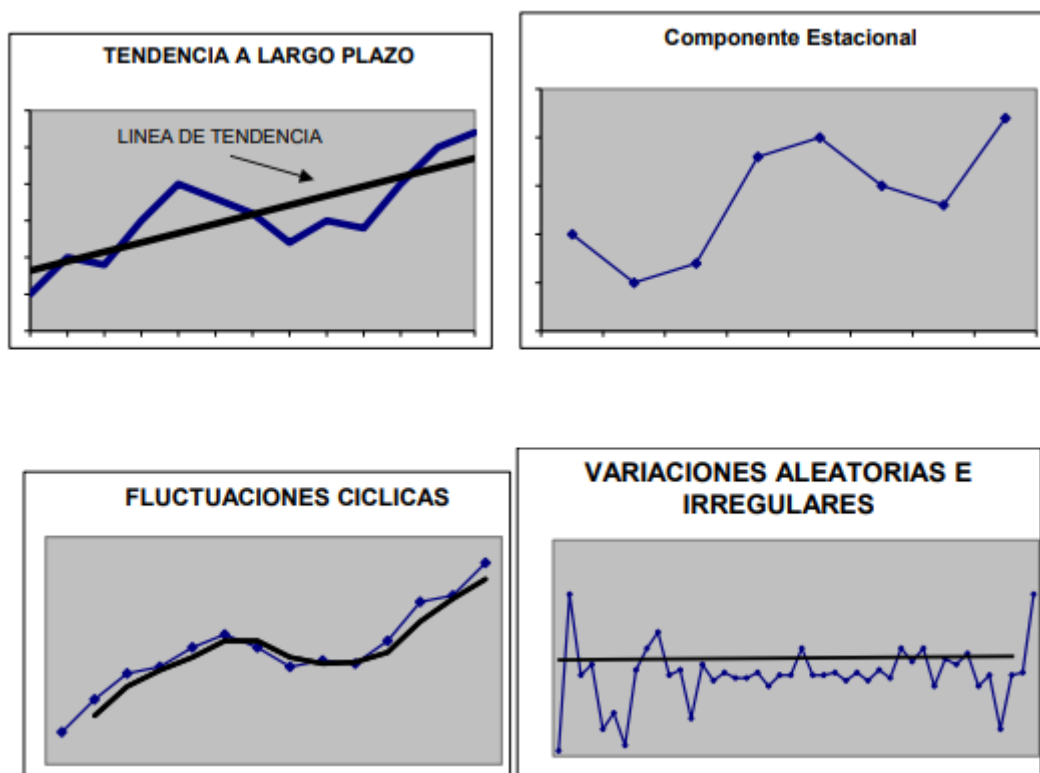
Estimación por el método de mínimos cuadrados



Se parte de considerar que la subpoblación de Y es normal, y que la suma de los cuadrados de las desviaciones de las observaciones respecto de la recta verdadera sea mínima

Debido a que se pretende encontrar el mejor modelo que ajuste a la nube de puntos se utiliza este método. • Se parte de considerar que la subpoblación de Y es normal, y que la suma de los cuadrados de las desviaciones de las observaciones

COMPONENTES DE UNA SERIE DE TIEMPO



El análisis de una serie de tiempo es un tema complicado; hay una variedad de opiniones en cuanto a como se tendrían que realizar los análisis. Uno de los enfoques de mayor aceptación es considerar una serie de tiempo como una combinación de 4 elementos, los cuales superpuestos y actuando en forma conjunta contribuyen a los cambios que se observan en un período de tiempo. Estos elementos son:

a)- Tendencia a largo plazo b)- Variación estacional c)- Variación cíclica d)- Variación aleatoria e irregular, impredecible

Estos componentes se aíslan y se ajustan utilizando algunos métodos y son:

a)- Tendencia a largo plazo: Es el movimiento de una serie de tiempo (creciente o decreciente) gradual en el tiempo de acuerdo a una curva

b)- Variación estacional: representa la tendencia de la serie de tiempo a variar hacia arriba y hacia abajo durante épocas específicas del año y más o menos con igual intensidad, pueden ser meses, bimestres o trimestres, además puede presentarse con movimientos periódicos por naturaleza. Es decir que esta componente se presenta cuando se trabaja con datos mensuales.

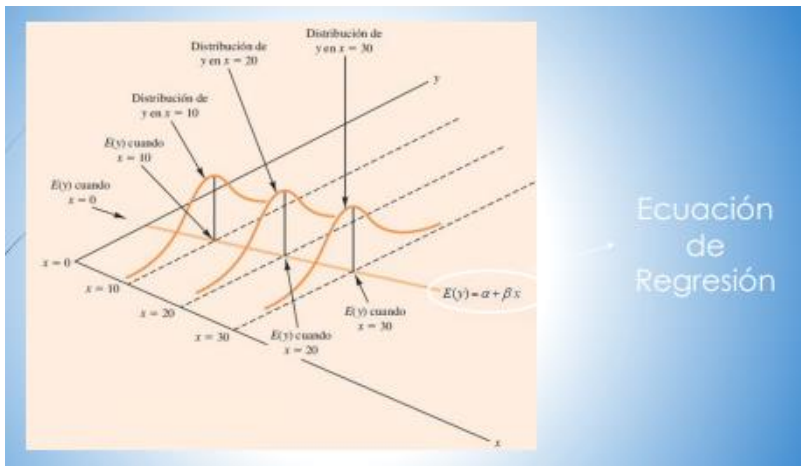
c)- Variación Cíclica: Los componentes cíclicos de una serie de tiempo tienden a subir o bajar según un patrón cíclico alrededor de la curva de la tendencia. Difieren de la componente anterior en que se extienden por períodos de tiempo más largos derivándose de causas diferentes.

d)- Variación aleatoria e irregular: Esta variación se añade a las anteriores. Se presentan de manera casual debido a efectos inexplicados como por ejemplo: -

Guerras - Inundaciones -

<i>Componente</i>	<i>Definición</i>	<i>Razón</i>	<i>Duración</i>
Tendencia	Patrón de movimientos ascendentes o descendentes, persistente a largo plazo.	Debido a cambios en la tecnología, riqueza, población.	Varios años.
Estacional	Fluctuaciones periódicas regulares que ocurren dentro de cada período de 12 meses, año a año.	Debido a condiciones de costumbres, tiempo, etc.	Dentro de los 12 meses.
Cíclica	Movimientos repetitivos ascendentes y descendentes mediante cuatro fases: -punto más alto: Prosperidad -contracción: Recesión -simas: Depresión -expansión: Recuperación.	Interactúan una serie de combinaciones de factores que influyen en la economía.	Generalmente de 2 a 10 años con diferente intensidad para cada ciclo completo.
Irregular	Fluctuaciones que existen en una serie luego de tomar en cuenta los efectos sistemáticos anteriores.	Se relacionan con acontecimientos imprevistos como huelgas, inundaciones, etc.,	Breves y no repetitivas.

REGRESION Y CORRELACION



Significado de los parámetros

α : intercepción de la línea de regresión con el eje Y.

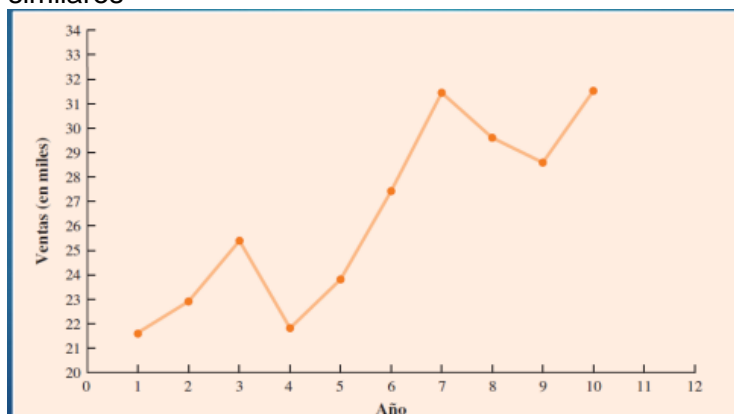
β : pendiente de la recta, proporción de cambio en la media de la distribución de probabilidad de Y por unidad de cambio de X.

TENDENCIA:

Aunque los datos de las series de tiempo presentan fluctuaciones aleatorias, también pueden mostrar cambios o movimientos graduales hacia valores relativamente mayores o menores durante un período.

Si una gráfica de serie de tiempo muestra este tipo de comportamiento, se dice que existe un patrón de tendencia.

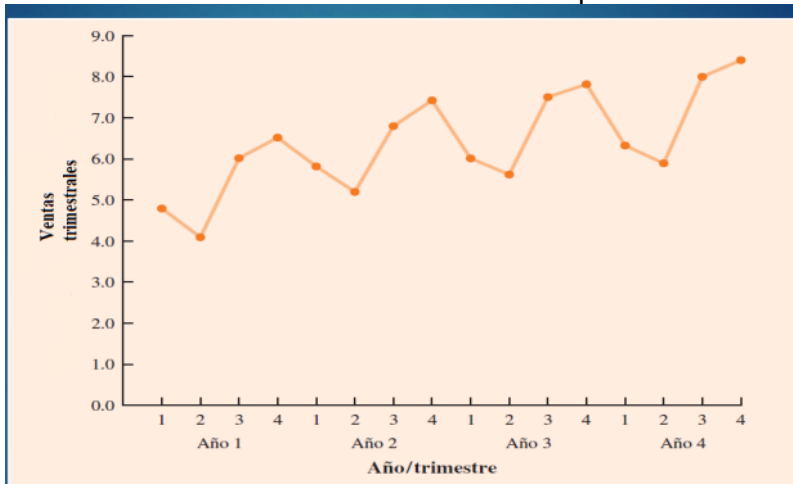
La tendencia, por lo general, es el resultado de factores a largo plazo, como el aumento o disminución de la población o la variación de sus características demográficas, la tecnología y/o preferencias de los consumidores, y cuestiones similares



ESTACIONALIDAD

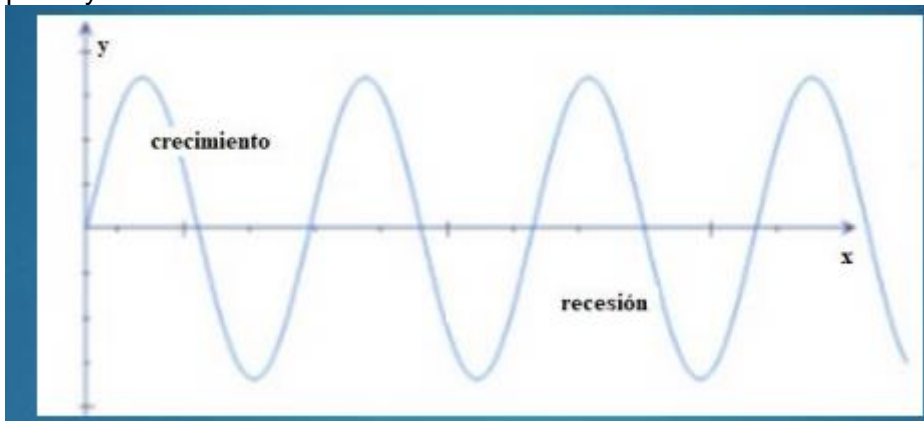
Los patrones estacionales son reconocidos al identificarse los mismos patrones de repetición en periodos sucesivos. El patrón de una gráfica de serie de tiempo que tiene un comportamiento repetitivo en un periodo de un año debido a la influencia estacional se llama patrón estacional

En general se considera que las variaciones estacionales son aquellas que se presentan en un período de un año, los datos de series de tiempo también pueden presentar patrones estacionales de menos de un año. Por ejemplo, el volumen de tráfico diario muestra en un día un comportamiento “estacional”, donde los valores máximos se presentan en las horas pico, un flujo moderado el resto del día y al comienzo de la noche, y un flujo ligero desde la medianoche hasta la madrugada. Lo mismo sucede con el tráfico de información por internet.



CICLICO

El patrón cíclico existe si la gráfica de la serie de tiempo muestra una secuencia de puntos que caen de manera alterna por arriba y debajo de la línea de tendencia por más de un año. Muchas series de tiempo económicas suelen mostrar un comportamiento cíclico. A menudo, el patrón cíclico se debe a ciclos multianuales de la economía. Por ejemplo, periodos de inflación moderada seguidos por periodos de inflación rápida y alta pueden dar lugar a que la serie de tiempo alterne hacia arriba y hacia abajo de la línea general de tendencia creciente (por ejemplo, una serie de tiempo sobre el costo de vivienda). No dependen del clima o factores sociales. Los ciclos económicos son extremadamente difíciles, si no imposibles de predecir. Como resultado, los efectos cíclicos a menudo se combinan con efectos de tendencia a largo plazo y se conocen como efecto de tendencia-cíclico



PROMEDIOS MOVILES:

Un promedio móvil se elabora sustituyendo cada valor de una serie por la media del mismo y alguno de los valores inmediatos anteriores y posteriores. “Un promedio móvil M, de K puntos en un momento dado del tiempo t, se calcula promediando K valores en secuencia de Y. El tiempo t se toma como el punto medio de estos intervalos

El objetivo: eliminar, tanto como sea posible, las variaciones indeseables de los datos, tratando de darle a la serie un aspecto más uniforme. El problema está en la elección apropiada del período para el promedio, esto depende de la naturaleza de los datos y el propósito que se persigue

SUAVIZACION EXPONENCIAL

la suavización exponencial es un tipo especial de promedio móvil, pero su naturaleza es muy diferente del cálculo del Promedio Móvil

$$S_i = \omega * Y_i + (1 - \omega) * S_{i-1}$$

S_i : Valor de la serie exponencialmente suavizada calculada en el período i

S_{i-1} : Valor de la serie exponencialmente suavizada ya calculada en el período i-1

Y_i : valor observado de la serie de tiempo en el período i ω : ponderación o coeficiente de suavización asignado en forma subjetiva ($0 < \omega < 1$)

Si solo se quiere suavizar una serie mediante la eliminación de variaciones cíclicas e irregulares que no se desean, debe seleccionarse un valor pequeño de ω (cercano a 0). Si se quieren hacer pronósticos se elegirá un valor grande de ω (cercano a 1)

Si el valor de ω es demasiado grande no se suavizarán adecuadamente las variaciones irregulares. Si ω es demasiado chico, el promedio móvil será insensible a las variaciones que se puedan efectuar en realidad

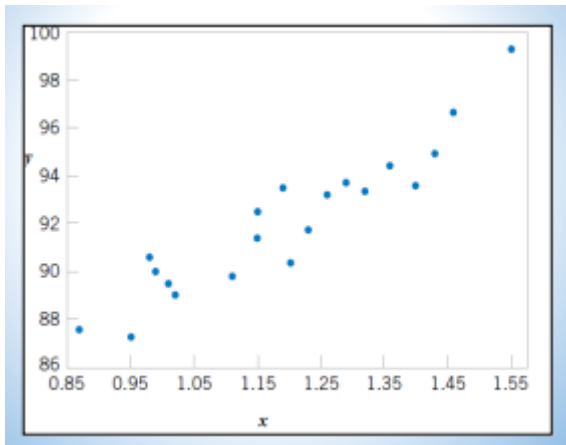
DESESTACIONALIZACION

Una serie de tiempo a la que se le han eliminado los efectos estacionales se conoce como serie de tiempo desestacionalizada, y al proceso de uso de los índices estacionales para eliminar los efectos estacionales de una serie de tiempo se le conoce como desestacionalizar la serie de tiempo

ANALISIS DE REGRESION

DISPERSIOGRAMA

Se comienza realizando el gráfico que permite visualizar los valores de las variables, es lo que se denomina Dispersiograma, ya que muestra la variabilidad o dispersión existente entre ambas variables



REGRESIÓN LINEAL SIMPLE • Es el análisis en el que se estudia la relación en la que interviene una variable independiente y una variable dependiente y que se

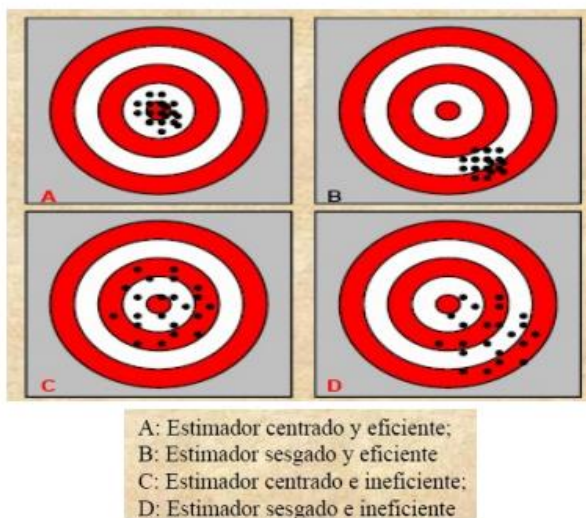
aproxima mediante una línea recta. • Al análisis en el que intervienen dos o más variables independientes se le llama análisis de regresión múltiple

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

α y β parámetros del modelo, que deberán ser estimados X_i variable independiente, fija, conocida, variable explicativa. ε término de error aleatorio

ESTIMADOR PUNTUAL

GRÁFICAMENTE



Es una función de valores observados (muestra) que no depende de ningún parámetro desconocido. Suponiendo que $\hat{\theta}$ es un estimador de θ , será considerado un “buen estimador” si cumple con las siguientes propiedades:

Insesgado: La esperanza del estimador coincide con el parámetro a estimar

$$E(\hat{\theta}) = \theta$$

Eficiente: Su varianza es lo menor posible

Consistente: Cuando el tamaño de la muestra crece arbitrariamente, se aproxima al valor del parámetro

INTERVALO DE CONFIANZA

El método de estimación por intervalos de confianza se basa en el valor de la estimación puntual pero considera además la distribución en el muestreo del estimador. Dada una variable aleatoria con distribución normal, podremos estimar con una cierta confianza un parámetro desconocido de su planteando la siguiente probabilidad

$$P(a \leq \hat{\theta} \leq b) = \text{confianza}$$

