

HERRAMIENTAS DE ESTADÍSTICA

Generalidades

- 1. Medidas Descriptivas**
- 2. Muestreo**
- 3. Diagrama de Flujo**
- 4. Hoja de Verificación o Inspección**
- 5. Diagrama de Pareto**
- 6. Diagrama Causa-Efecto**
- 7. Histograma**
- 8. Gráfica de Control**

GENERALIDADES

Se describen a continuación algunas herramientas y técnicas que han sido desarrolladas para el mejoramiento de la calidad y el control de los procesos de trabajo. La aplicación de algunas de éstas, dependerá de las necesidades de cada área y la que mejor se adapte a los registros de calidad que se generan en el desarrollo de sus actividades.

Herramientas para datos numéricos

Las decisiones para el mejoramiento de la calidad están basadas en datos numéricos para toda actividad. Incluso lo cualitativo puede ser agrupado y clasificado numéricamente para analizarlo de manera más efectiva, las decisiones en cuanto a diferencias, tendencias y cambios en datos están basadas en la interpretación estadística.

MEDIDAS DESCRIPTIVAS

En todo conjunto de datos es posible estadísticamente diferenciar entre dos tipos de medidas:

- Las medidas de tendencia central
- Las medidas de dispersión

A éstas se les llama **Medidas descriptivas**, ya que únicamente se utilizan **para describir y analizar** un grupo de datos, sin llegar a obtener conclusiones o inferencias.

Una medida de tendencia central es el dato al cual tienden a agruparse los demás datos de un conjunto.

Entre las principales medidas de tendencia central tenemos:

- Media Aritmética (promedio)
- Mediana
- Moda

A continuación se describe y ejemplifica cada una de ellas:

- **Media Aritmética:** Es el promedio aritmético de un conjunto de datos. Su representación es:

\bar{X} (testada ó X barra) y la ecuación para calcular es:

$$\bar{X} = \frac{\sum x}{n}$$

- Σ Suma de todos los datos
- x Todos y cada uno de los datos.
- n Número total de datos

- **Mediana:** Es el dato que divide en dos partes iguales a un conjunto de datos que han sido **ordenados** previamente.

De acuerdo con esta definición:

- Si el número de datos es impar se elige como mediana al dato central.
- Si el número de datos es par se calcula la **X** de los dos datos centrales y esa es la mediana. El símbolo para representar la mediana es "**M**".

- **Moda:** Es el dato que más se repite (el de mayor frecuencia). Puede suceder que no exista la moda, que sea única o que sea múltiple. El símbolo que representa la Moda es "m".

Medidas de dispersión: Se entiende por dispersión o variación al grado en que los datos de muestra o población tienden a extenderse alrededor de un valor medio.

Las medidas de dispersión más usadas son:

- Rango
- Desviación típica o estándar.

A continuación se describe cada una de ellas.

- **Rango:** Es la diferencia que existe entre el dato mayor y el menor de un conjunto de datos, su representación es "R".
 $R = X_{\text{máx.}} - X_{\text{mín.}}$

MUESTREO

El muestreo es el proceso de tomar una proporción o parte de un **universo** de elementos, con la finalidad de analizar en dichos elementos, características sujetas a estudio o fenómenos factibles de observación y en base al análisis de la **muestra** o proporción tomada obtener conclusiones que se refieran no sólo a **la muestra** sino a todo el **universo**.

Para fines estadísticos, el universo puede considerarse **finito** o **infinito**. Se considera finito si el número de elementos que lo constituyen es menor a 500,000 e infinito si es igual o mayor a este número.

Muestra: es la parte del universo que presenta características similares a aquél.

La utilidad de emplear muestras se ejemplifica con el ama de casa que quiere saber si ha condimentado suficientemente la sopa. Para lo cual toma una cucharada de sopa, la prueba y en base a esta pequeña muestra obtiene conclusiones referentes no sólo a la muestra que tomo, sino a toda la sopa preparada (**universo**).

Siempre que hagamos la elección de una muestra, debemos tener cuidado de que ésta reúna las siguientes características:

- Que sea **suficiente**: es decir que la cantidad de elementos seleccionados sea el que se requiere para que el nivel de confiabilidad sea el que se ha establecido previamente.
- Que sea **representativa**: esto quiere decir que los elementos seleccionados deberán presentar características similares a las de la población o universo.

Al utilizar muestras en lugar de universos tenemos grandes ventajas, algunas de las más importantes son:

- El **costo** se reduce, pues los gastos serán únicamente los ocasionados por una parte del universo (muestra tomada) y no por la totalidad de él.
- Si la muestra es representativa, las deducciones resultantes sobre el universo serán **confiables**.
- Como solamente se estudia una parte del universo, la información obtenida se realiza en **menor tiempo**.

¿ Cómo obtener el tamaño de la muestra a utilizar?

1. Determinamos el nivel de confianza con el cual vamos a trabajar y buscamos el valor de z asociado a dicho nivel de confianza, un nivel de confianza igual o mayor al 92% es aceptable estadísticamente.

2. Evaluamos la probabilidad a favor de que suceda un evento o situación esperada (esta probabilidad se le denomina p).

3. Evaluamos la probabilidad en contra de que suceda en un evento o situación esperada (a esta probabilidad se le denomina q).

4. Determinamos el error máximo para el nivel de precisión que vayamos a permitir en los resultados (error máximo de estimación), comúnmente se trabaja con errores de estimación entre el 2% y el 6%, ya que la validez de la información se reduce demasiado para valores mayores del 6%.

· Determinamos el tamaño de la población o universo.

5. Se elige la fórmula a utilizar para calcular el tamaño de la muestra; dependiendo de si la población o universo sujeto a estudio se va a considerar infinito ó finito. (Una población o universo se considera infinito si el número de elementos de los que consta es igual o mayor a 500,000 y es considerado finito si el número de elementos es menor a esta cantidad).

Para poblaciones o universos considerados **infinitos** la fórmula es :

$$\boxed{n = \frac{z^2 pq}{e^2}} \quad \text{ó} \quad \boxed{n = \frac{z^2 p(1-p)}{e^2}} \quad \text{si } p = q = 50\% \text{ entonces: } n = \left[\frac{z}{2e} \right]^2$$

Para poblaciones o universos considerados finitos la fórmula es:

$$n = \frac{z^2 N pq}{e^2 (N-1) + z^2 pq}$$

En donde:

n = Tamaño de la muestra.

z = Valor en Tablas asociado a la desviación estándar para un nivel de confianza deseado.

p = Probabilidad a favor de que suceda un evento o situación esperada.

e = Error de estimación.

N = Tamaño de la población o universo a estudiar (número de elementos).

3. **DIAGRAMA DE FLUJO**

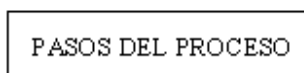
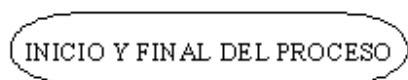
· ¿Qué es?

Es una representación gráfica que muestra todos los pasos de un proceso y la relación que guardan unos con otros.

· ¿Cómo se elabora?

Esta herramienta se elabora utilizando símbolos fáciles de reconocer con la finalidad de representar el tipo de operación realizada.

Los símbolos son:



Analizando cuidadosamente el diagrama de flujo de un proceso, comúnmente descubrimos vacíos, los cuales son fuentes potenciales de problemas.

Cuando se vaya a elaborar un diagrama de flujo se recomienda que la(s) persona(s) que tenga(n) mayor conocimiento del proceso:

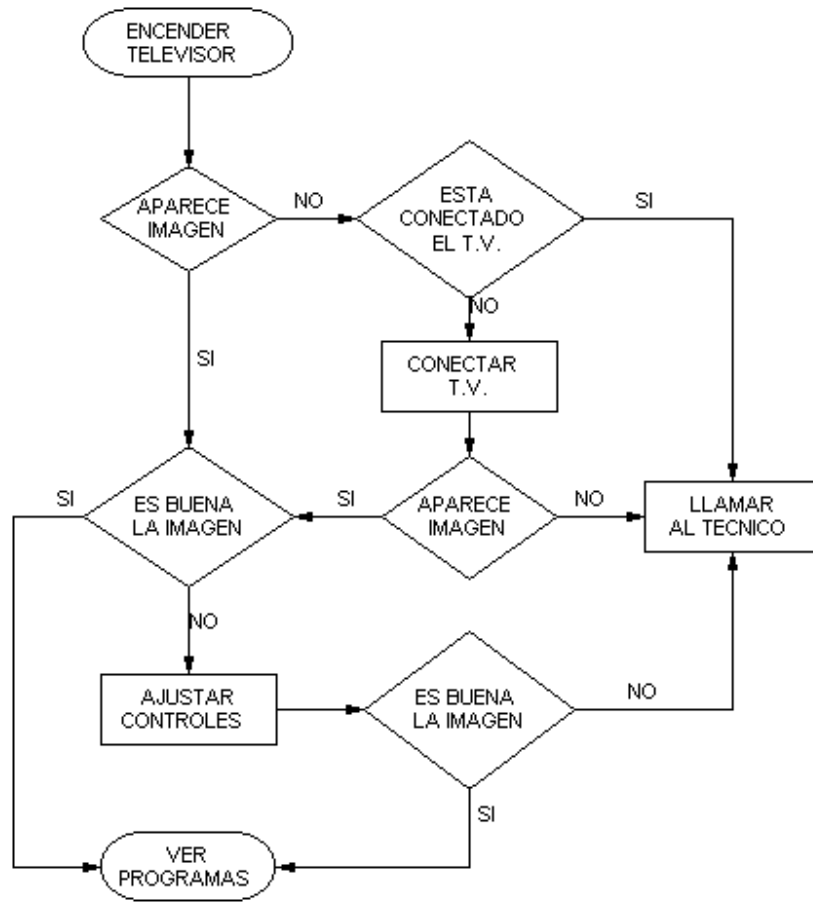
- Trace(n) un diagrama de flujo del proceso actual.
- Trace(n) un diagrama de flujo, indicando los pasos que éste debería seguir si todo trabajara en forma correcta.
- Compare(n) los dos diagramas y encuentre(n) las diferencias, ya que es ahí donde radican los problemas.

· Utilización.

Se utiliza cuando se hace necesario identificar la trayectoria actual y la ideal que sigue un producto o servicio con el fin de identificar desviaciones (problemas).

Veamos un ejemplo sencillo de diagrama de flujo en la vida diaria de cada uno de nosotros.

Encendido de un T.V.



4. HOJA DE VERIFICACIÓN O INSPECCIÓN

· ¿Qué son?

Son formatos sencillos los cuales nos ayudan a recolectar datos.

· ¿Cómo se elaboran?

Para su elaboración se requiere primeramente que dicho formato conteste a la pregunta:

· ¿Qué tan frecuente ocurre el evento sujeto a estudio u observación?

Son requisitos indispensables para la elaboración de las hojas de verificación o inspección los siguientes:

- Todas las personas involucradas en la recolección de los datos deben estar de acuerdo sobre el evento que va a ser observado.
- Tomar la decisión del período de tiempo durante el cual serán recolectados del dato (horas, semanas, meses, etc.)
- Diseñar una forma clara y fácil de usar para el registro de los datos.
- Asegurarse de que al obtener los datos este trabajo se haga de una manera consistente y honesta.
- Asegurarse de que las muestras u observaciones sean tomadas al azar.

· Utilización:

Esta herramienta se utiliza cuando se necesita reunir datos basados en la observación con la finalidad de detectar tendencias y comenzar a analizar los problemas subyacentes de éstas; es decir, aquí comienza el proceso de convertir simples opiniones o suposiciones de hechos.

La utilización de las hojas de verificación es el punto lógico de inicio para el análisis de solución de problemas.

Ejemplo de servicio: se realizó un muestreo de los servicios de cajero automático durante un mes en el banco "X" y los resultados de los problemas que se observaron fueron los siguientes:

SERVICIOS DE CAJEROS AUTOMÁTICOS DEL BANCO "X"

PROBLEMAS	FEBRERO (SEMANAS)				TOTAL
	1	2	3	4	
No dio el dinero solicitado	II	IIII	III	IIII-II	18
No tiene dinero disponible	IIII IIIII	IIII IIIII IIIII	IIII IIIII IIIII	IIII IIIII IIIII IIIII	72
No estaba en red con el banco de la tarjeta usada	III	IIII	II	IIII	14
No estaba en servicio	IIII IIIII IIIII	IIII IIIII IIIII IIIII	IIII IIIII IIIII IIIII	IIII IIIII IIIII IIIII	86
No devolvió la tarjeta	I	III	I	IIII	10
					200

Ejemplo de administración / vida diaria: en el lapso de un mes se registraron en el Departamento de Urgencias del hospital "Z" los casos de lesiones que se atendieron, siendo el resultado de estas observaciones el que se muestra en la siguiente hoja de verificación o inspección:

5. DIAGRAMA DE PARETO

· ¿Qué es?

Gráfica que representa de una forma ordenada (de mayor a menor) la ocurrencia de factores sujetos a estudios, tales como: fallas, errores, defectos, quejas, necesidades, etc.

· ¿Cómo se construye o elabora?

Las reglas para la elaboración de un diagrama de Pareto se darán a partir del ejemplo del servicio de cajeros automáticos del banco "X" del cual se hizo la hoja de verificación.

- Reglas o pasos para la construcción de un diagrama de Pareto.
- Seleccionar el factor a analizar. En este caso se analizó el servicio de cajeros automáticos.
- Ordenar los problemas, fallas, quejas, necesidades, etc., de mayor a menor incidencia o frecuencia.

Tipo de Problema	Frecuencia o incidencia
· No estaba en servicio	86
· No tenía dinero disponible	72
· No dio completo el dinero solicitado	18
· No estaba en red con el banco de la tarjeta usada.	14
· No devolvió la tarjeta	10
	200

- Obtener la frecuencia acumulada.

Tipo de problema	Frecuencia	Frecuencia acumulada
· No estaba en servicio	86	86
· No tenía dinero disponible	72	158
· No dio completo el dinero solicitado.	18	176
· No estaba en red con el banco de la tarjeta usada.	14	190
· No devolvió la tarjeta	10	200

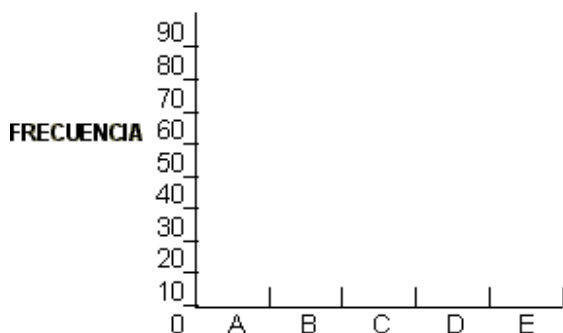
- Calcular el porcentaje de la frecuencia acumulada.

Código	Tipo de Problema	Frec.	Frec. Acum	% de frec. Acum..
A	No estaba en servicio	86	86	43.0
B	No tenía dinero disponible.	72	158	79.0
C	No dio completo el dinero solicitado	18	176	88.0
D	No estaba en red con banco de tarjeta usada.	14	190	95.0
E	No devolvió la tarjeta	10	200	100.0
		200		

- Trazar un eje horizontal y un eje vertical:

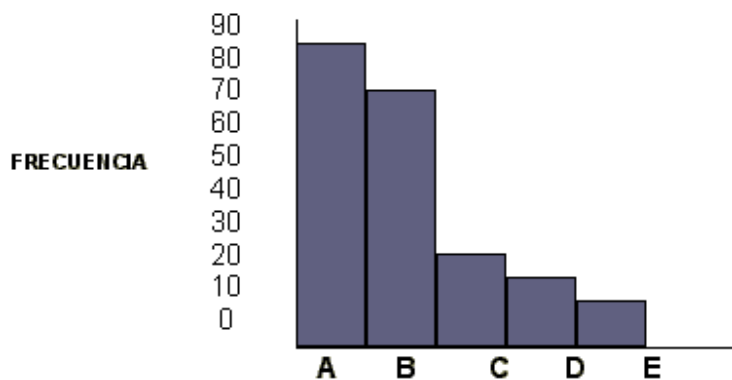
En el eje horizontal seleccionar un intervalo adecuado para representar cada tipo de problema, defecto, queja, necesidad, etc., identificando cada uno de ellos en cada intervalo seleccionado.

En el eje vertical elegir una escala adecuada (fácil de leer) en números enteros para representar la frecuencia.



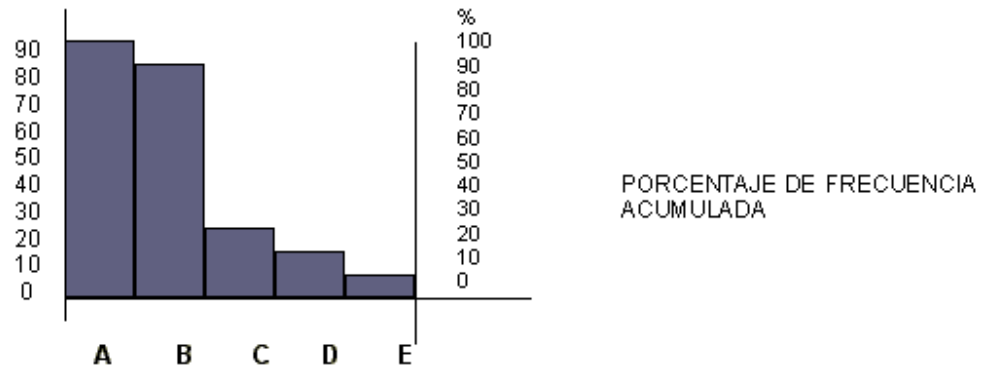
TIPO DE PROBLEMA

- Construir barras correspondientes a cada tipo de problema, queja, necesidad, etc., y a su frecuencia.

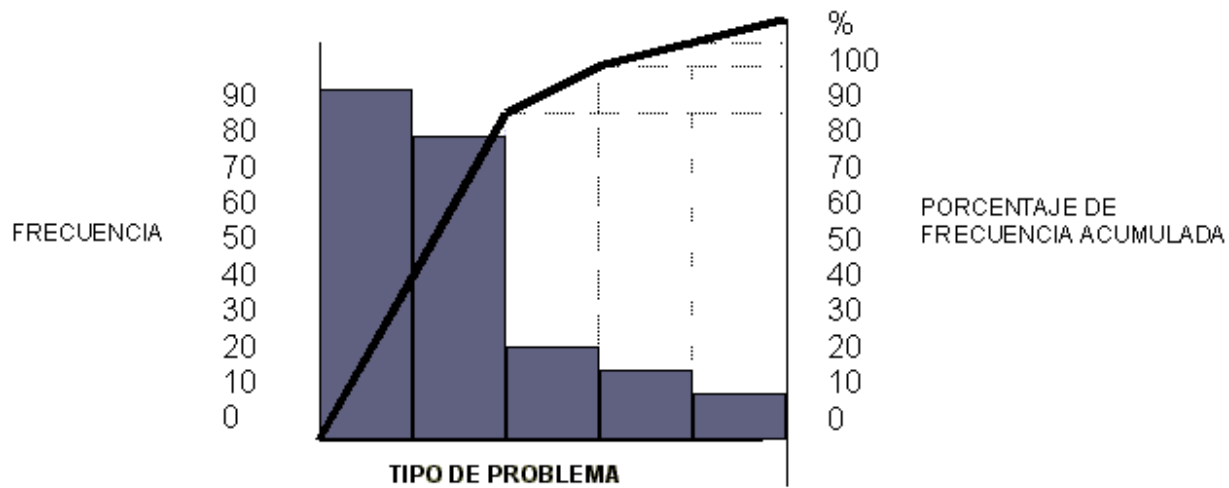


TIPO DE PROBLEMA

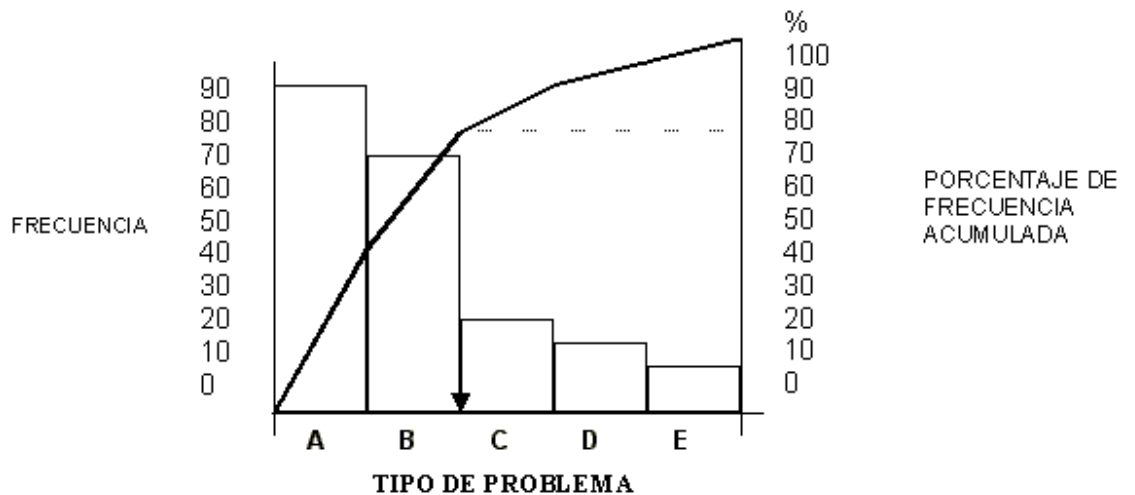
- Trazar otro eje vertical sobre la pared donde termina la última de las barras del paso anterior y elegir una escala adecuada para graficar el porcentaje de la frecuencia acumulada (la escala de este nuevo eje puede ser igual o diferente a la de la frecuencia).



- Trazar la línea de porcentaje de frecuencias acumuladas.



- Determinar los factores vitales y triviales (80-20). Para esto se traza una línea horizontal que inicie en el 80% del eje vertical derecho, de manera que toque la línea de frecuencia acumulada y de este punto se baja una vertical hasta tocar el eje horizontal; los defectos, problemas, quejas, etc. que hayan quedado a la izquierda de esta línea son los factores vitales, el resto los triviales.



Pareto acostumbrada tomar el 80-20 pero se puede analizar otro porcentaje, por ejemplo 75-25 ó 70-30. etc.

- Utilización
- Determinar qué problemas resolver y en qué orden.
- Separar factores vitales de triviales.
- Determinar los factores con mayor área de oportunidad.
- Confirmar el efecto de una mejora realizada al proceso.
- Conocer si un proceso está o no bajo control.

- Concepto del 80-20:

El 80% de un problema está generado por sólo el 20% de las causas.

Es muy importante utilizar en el diagrama de Pareto el sentido común ya que no necesariamente los eventos más frecuentes son los más importantes, por ejemplo: dos accidentes fatales requieren más atención que 100 cortaduras de dedos.

Ejercicio:

Construir el diagrama de Pareto del ejemplo de las lesiones registradas en el Departamento de Urgencias del hospital "Z" que se vio en el apartado 4 (hoja de verificación o inspección).

6. DIAGRAMA CAUSA-EFECTO

- ¿Qué es?

Es una herramienta en la que se muestran todas las posibles causas que origina un determinado problema, el cual está siendo analizado.

Un diagrama de causa-efecto bien detallado, comúnmente toma la forma de un "ESQUELETO DE PESCADO", por lo que también recibe este nombre. Este diagrama fue desarrollado por el japonés Kaoru Ishikawa en el año de 1953.

- ¿Cómo se construye?

- Decidir el problema a analizar.
- Generar un listado de todas las posibles causas que estén incidiendo en el problema que está siendo analizado.

El listado se obtiene a través de una técnica que se conoce como LLUVIA DE IDEAS, para lo cual se requiere de la experiencia de la gente que conoce de la situación problemática y de todas las posibles causas que puedan estar afectando a un problema en particular.

Lluvia de ideas

La lluvia de ideas tiene dos modalidades.

Estructurada: En esta modalidad cada persona del equipo de trabajo debe dar alguna idea conforme le toca el turno de participar, si no se tiene ninguna idea debe esperar su turno en la siguiente vuelta. (Esta modalidad fuerza a participar a las personas tímidas, pero crea cierta presión a participar a todos).

No estructurada: Esta modalidad deja en libertad para que cada miembro del equipo de trabajo aporte sus ideas tan pronto le vienen a la mente. (Se crea un ambiente más relajado, pero el riesgo es que únicamente participen los más extrovertidos).

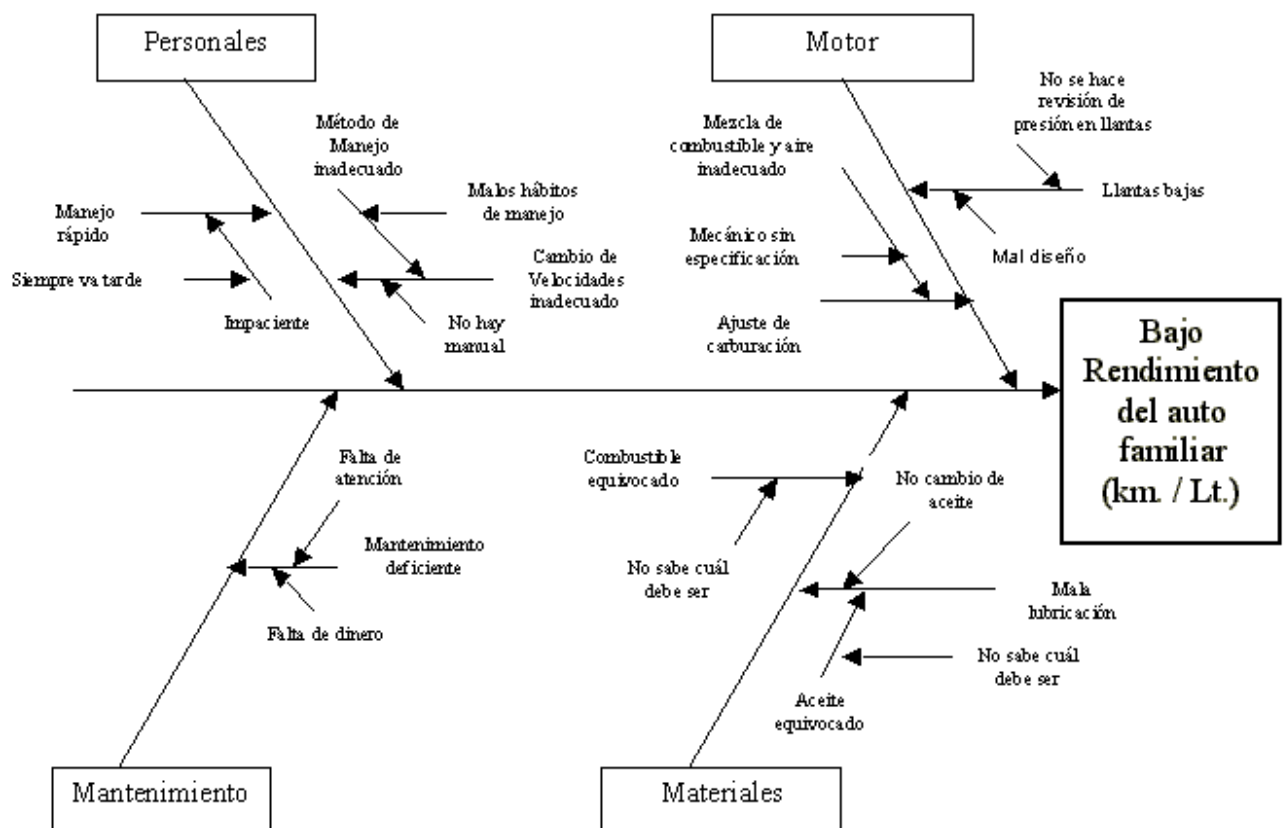
- Al momento de expresar las ideas, nunca se deben criticar.
- Se escribirá cada idea en un pizarrón o rotafolio para que todos puedan leerlas (esto evita malos entendidos y ayuda a crear nuevas ideas).
- Anotar las palabras del que aporta la idea, no interpretarlas a su manera.
- No buscar culpables cuando sugieren ideas de causas de problemas.

- Escribir en un recuadro, al extremo derecho de una flecha dibujada como base del diagrama, una frase que describa o identifique el problema a analizar.

- Dibujar flechas inclinadas que apunten a la principal, una para cada grupo de grandes causas, en las cuales se van a agrupar las causas (ideas) que resultarán de la "lluvia de ideas". Generalmente se utilizan 4 grupos de causas, en manufactura se acostumbra que estas sean: mano de obra, método, maquinaria y materiales; en problemas administrativos, de servicios de la vida diaria no existe un patrón, pudiendo ser: presupuestales, de personal, políticas, administrativas, métodos y procedimientos, normativas, etc.

- Anotar sobre las ramas de grupos de causas los factores o causas más pequeñas o secundarias.

Ejemplo de la vida diaria sobre las causas por las cuales podría suceder que el rendimiento del auto familiar es bajo.



• Utilización

Se utiliza cuando se requiera analizar y mostrar todas las causas posibles de un problema para su posterior solicitud.

7. HISTOGRAMA

- ¿Qué es?

Es la representación gráfica por medio de barras, en forma ordenada de las veces que ocurren las variaciones en un proceso.

- ¿Cómo se construye?

Los pasos para su construcción se darán a partir del siguiente ejemplo para que se comprendan mejor.

Supongamos que los siguientes datos son las calificaciones de un grupo de alumnos de la Universidad "Y" en la materia de estadística; las calificaciones están dadas a un decimal.

5.1	6.3	7.6	8.0	9.1	10.0	8.4	7.8
7.2	8.4	4.9	8.7	6.9	7.0	6.5	10.0
7.8	6.7	7.8	9.0	7.7	7.7	5.8	6.7
9.3	7.9	6.2	7.9	6.8	5.4	7.7	9.0
8.5	9.2	6.3	8.2	7.3	7.0	8.4	7.3

Construyamos el histograma correspondiente:

- **Paso 1:** Obtener el rango de los datos de la tabla

$$R = X_{\text{máx}} - X_{\text{min}}$$

$$R = 10.0 - 4.9 = 5.1$$

Recordemos que el rango se obtiene de restar el dato mayor menos el dato menor.

- **Paso 2:** Contar el número de datos (n)

$$n = 40 \text{ (tamaño de la muestra)}$$

- **Paso 3:** Decidir el número de grupos, clases o categorías que se van a usar (Estadísticamente se recomienda que sean entre 5 y 8 si $n > 100$).

K = número de grupos, categorías o clases.

Supongamos que decidimos usar una K=7 para este histograma.

Nota: Se recomienda una K entre 7 y 12 si $n > 100$ pero menor a 250.

Se recomienda una K entre 12 y 20 si $n > 250$

- **Paso 4:** Determinar el tamaño o ancho de cada categoría, grupo o clase del histograma (es decir el ancho de cada una de las barras).

La ecuación para obtener esto es la siguiente :

$L = \frac{R}{K}$ L= longitud o ancho de cada categoría

K R=rango

K= número de categorías o clases

$$L = \frac{5.1}{7} = 0.728$$

7

Para fines prácticos se toma un decimal más de los que constan los datos del problema en cuestión, es decir, para este ejemplo:

$$L = 0.73$$

- **Paso 5:** Determinar los límites de cada clase o categoría.

Los límites se simbolizan con la letra "c"

El procedimiento es el siguiente:

El primer límite se toma idéntico al dato más pequeño, es decir en este ejemplo es 4.9, el segundo límite resulta de sumar "L" al primer límite y así sucesivamente hasta igualar o sobrepasar el dato mayor.

$$C1 = 4.9$$

$$C2 = 4.9 + 0.73 = 5.63$$

$$C3 = 5.63 + 0.73 = 6.36$$

$$C4 = 6.36 + 0.73 = 7.09$$

$$C5 = 7.09 + 0.73 = 7.82$$

$$C6 = 7.82 + 0.73 = 8.55$$

$$C7 = 8.55 + 0.73 = 9.28$$

$$C8 = 9.28 + 0.73 = 10.01$$

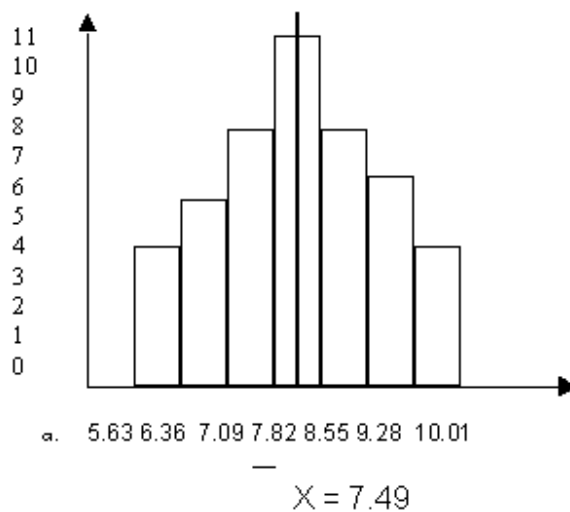
- **Paso 6:** Construir la tabla de frecuencias con los límites de clase encontrados en el paso 5, tomando para formar cada clase 2 límites consecutivos (C1,C2, C2-C3, C3-C4, etc.) y encontrar el número de datos que caen en cada clase o intervalo.

Clases o categorías	Frecuencia
4.90-5.63	3
5.63-6.36	4
6.36-7.09	7
7.09-7.82	11
7.82-8.55	7
8.55-9.28	5
9.28-10.01	3
TOTAL	40

• **Paso 7:** Construir el histograma.

Para determinar la altura de cada barra se selecciona una longitud adecuada, ésta depende de la frecuencia de cada clase o intervalo.

Se traza un eje horizontal en el cual se van a localizar los límites de clase.



• **Utilización:**

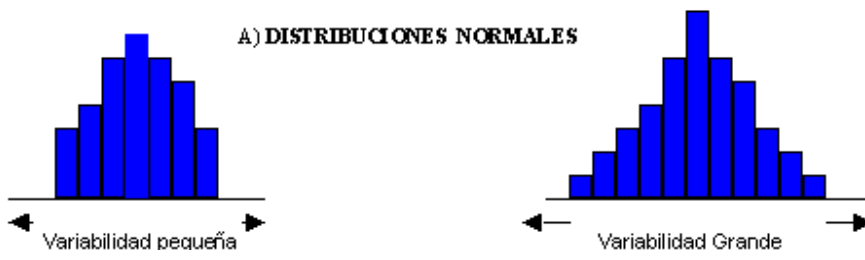
- Se usa cuando se quiere conocer la forma, localización y dispersión de los datos.
- Cuando se quiere confirmar los efectos de las mejoras realizadas a los procesos.

• **Patrones de Histogramas:**

Dependiendo de la distribución de los datos y su dispersión, pueden resultar diferentes tipos de patrones al construir el histograma correspondiente; en otras palabras, podemos decir que la forma o patrón de un histograma depende del comportamiento de los datos a través de los procesos.

Los principales patrones son los que se presentan a continuación:

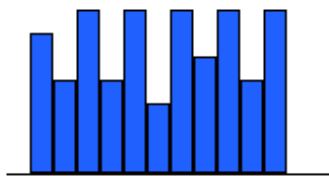
A) DISTRIBUCIONES NORMALES



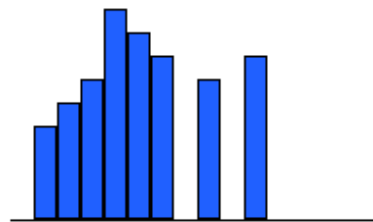
B) DISTRIBUCIONES SESGADAS



E) TIPO SERRUCHO



F) DATOS AISLADOS

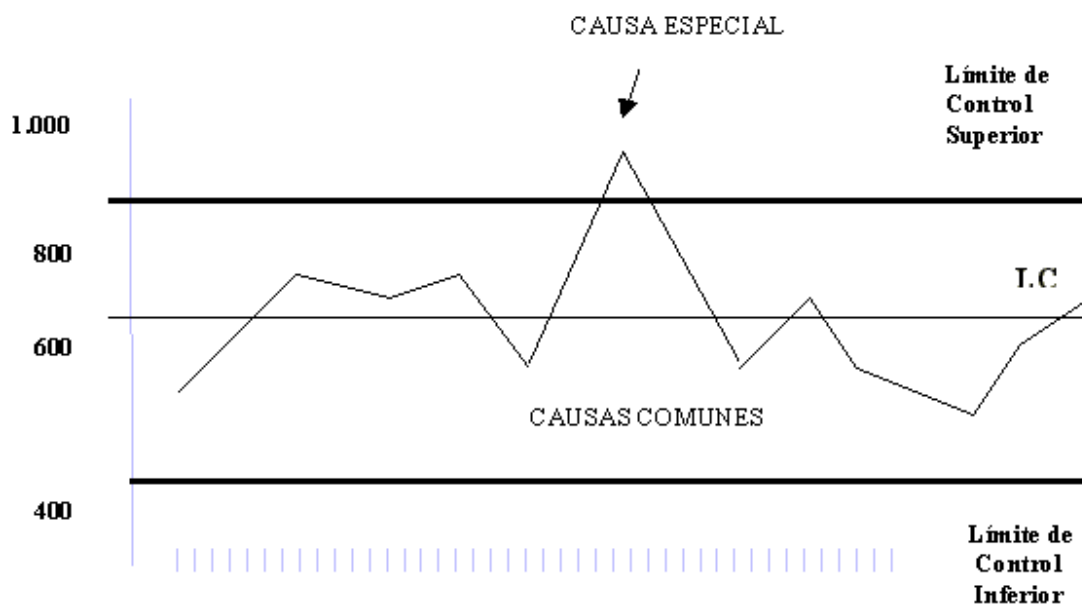


8. **GRAFICA DE CONTROL**

El histograma es una buena herramienta estadística pero sólo es una FOTOGRAFIA del proceso, en tanto que la gráfica de control es una PELÍCULA del proceso. Nos muestra como la media y los rangos cambian en el tiempo. Estos cambios reflejan como está cambiando el sistema completo.

El propósito de las gráficas de control (hay de varios tipos) es minimizar la posibilidad de hacer dos tipos de errores: 1) actuar como si algo extraordinario ha pasado cuando en verdad lo que sucedió es normal o esperado y, 2) dejar de actuar por pensar que lo que sucede es común cuando en verdad es extraordinario. Los estadísticos le llaman a estos errores de Tipo I y de Tipo II, respectivamente.

También se puede pensar en que estamos sobrecontrolando o subcontrolando un proceso. Esto es importante pues a veces sobrereaccionamos ante una variación normal del proceso o no actuamos por creer que la variación es normal.



En este tipo de gráfica se incluye un gran promedio de promedios o Límite Central (LC) así como un Límite Superior de Control (LSC). Todas las mediciones que se ubiquen dentro de estos límites superior e inferior se dice son variaciones inherentes al proceso o que se deben a causas comunes. En cambio, si alguna medición se ubica fuera de estos límites se debe a causas especiales.

Para construir una gráfica de control debemos seguir los siguientes pasos:

- Determinar el promedio de promedios = LC
- Calcular la X para cada subgrupo de los datos recolectados
- Calcular X es decir el promedio de promedios de los subgrupos

$$\bar{\bar{X}} = \frac{X_1 + X_2 + X_3 + \dots + X_k}{k} \quad k = \text{Número de subgrupos}$$

- Determinar el promedio de rangos \bar{R} .
- Calcular el R para cada subgrupo de los datos recolectados
- Calcular \bar{R} es decir el promedio de los rangos de cada subgrupo

$$\bar{R} = \frac{R_1 + R_2 + R_3 + \dots + R_k}{k} \quad k = \text{Número de subgrupos}$$

- Determinar el límite de control superior

$$LSC_{\bar{x}} = \bar{X} + A^2 \bar{R} \quad \text{ó} \quad LSC_{\bar{x}} = LC + A^2 \bar{R}$$

$$LSC_R = D_4 \bar{R}$$

- Determinar el límite inferior de control

$$LIC_{\bar{x}} = \bar{X} - A^2 \bar{R} \quad \text{ó} \quad LIC_{\bar{x}} = LC - A^2 \bar{R}$$

$$LIC_R = D_3 \bar{R}$$