



MEDIR

Análisis de Procesos y Toma de Datos



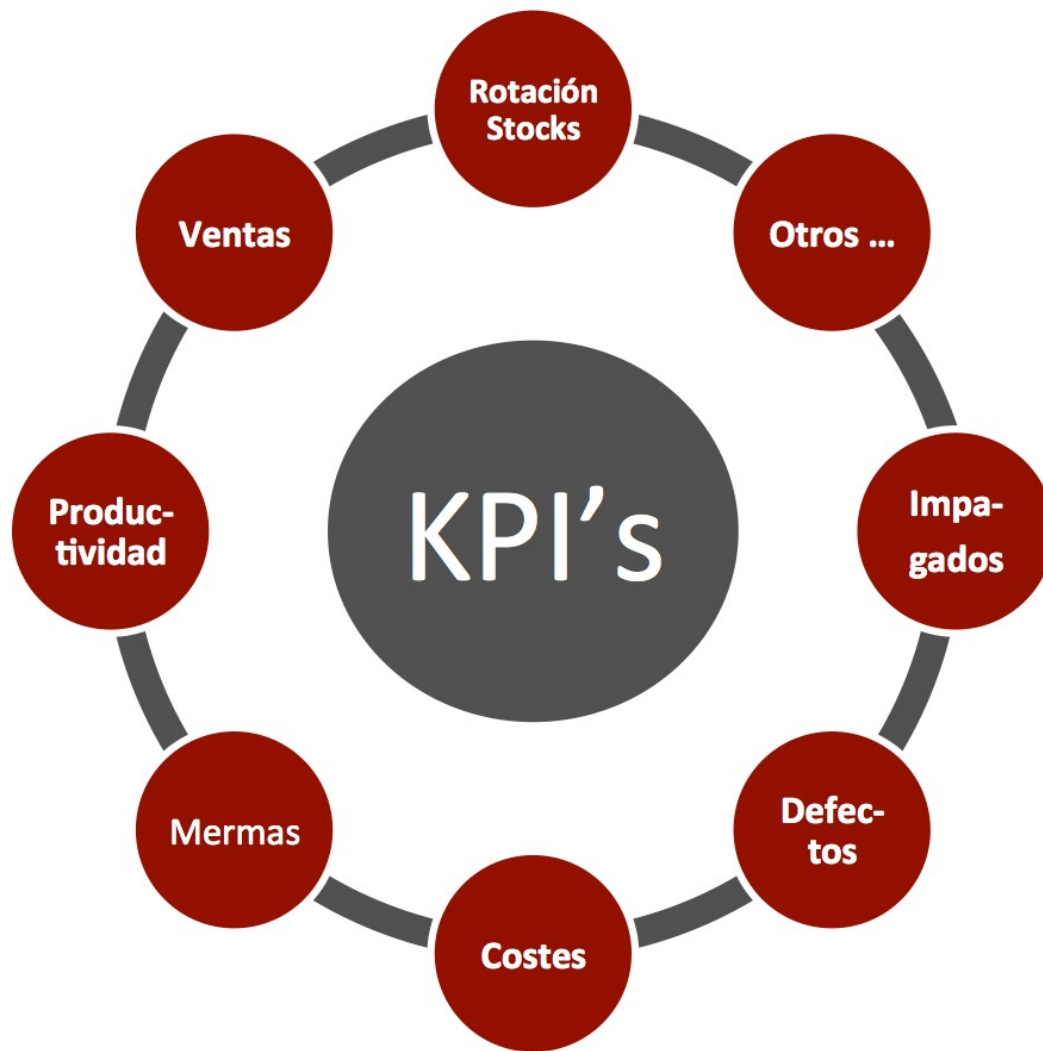
Paradigmas acerca de la medición

1. La medición precede al castigo
2. No hay tiempo para medir
3. Medir es difícil
4. Hay cosas imposibles de medir
5. Es más costoso medir que hacer



Paradigmas acerca de la medición

1. La medición precede al castigo
2. No hay tiempo
3. Es más fácil
4. Hay que medir
5. Es más costoso medir que hacer



Medir la variable más representativa o la que mejor tipifique el o los aspectos vitales del fenómeno.

¿Por qué medir?

Por obligación

Para compararse con la competencia

Para mejorar

Para buscar eficiencia

Para aprender

Para evaluar

Para motivar

Para presupuestar

Para controlar

Para liderar



MEDIR

Las mediciones describen las características del proceso bajo estudio, particularmente su capacidad.



CTC & CTB

En la etapa de medición buscamos identificar todas las variables que podrían estar impactando los requerimientos que son críticos para el cliente y los que son críticos para el negocio.

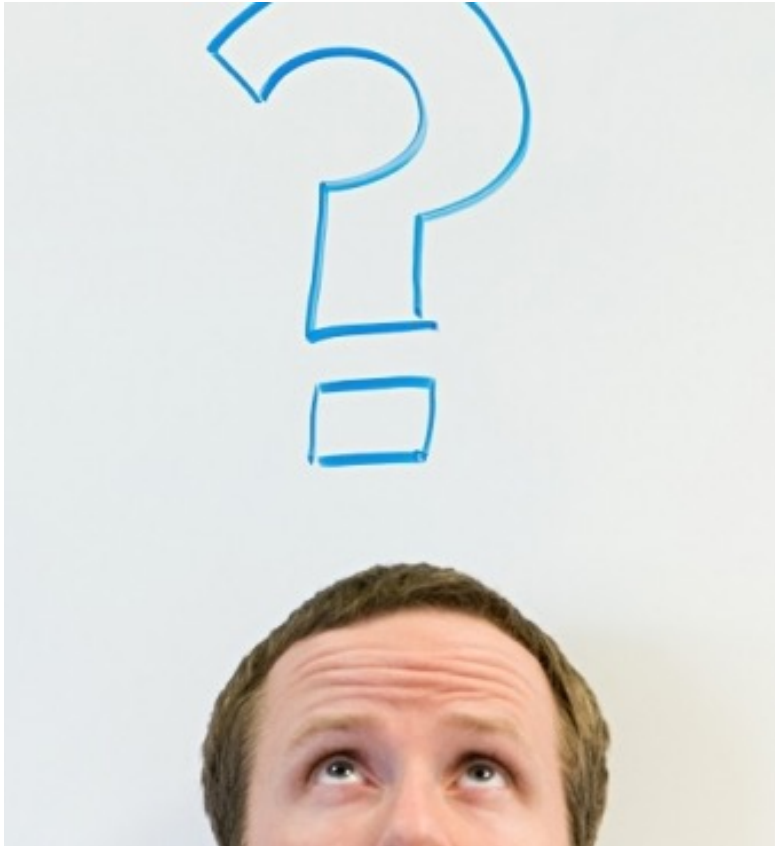


Recordar Herramientas que ya se vieron en DN-0496

1. Mapeo de Procesos
2. Diagramas de Flujo
3. SIPOC
4. Diagrama de Causa y Efecto
5. Pensamiento Lean
6. Valor Agregado
7. Muda
8. 8 tipos de desperdicio



Herramientas que ya se vieron en DN-0496



9. Flujo de Valor
10. Sistemas Pull
11. Kanban
12. Takt Time
13. Throughput
14. Value Stream Map
15. Requerimientos CTX
16. Kaizen



Conceptos Básicos de la Medición

Observe primero, luego mida

Observe el proceso. Converse e interactúe con las personas que trabajan en el proceso, aprenderá mucho de esta forma, qué es importante y debe ser medido en detalle, qué no es importante.

Recuerde que la variabilidad es el enemigo y este se puede detectar cuando se observa el proceso en detalle.



Mida con un propósito

No caiga en la trampa de medir por medir.

- **Mida eficiencia y efectividad.** Determine el resultado de un proceso desde el punto de vista de las salidas que este proceso logra y los recursos que utiliza para lograrlo. Incluye las mediciones de calidad, desperdicios, producción, etc. Este tipo de medida le permite al equipo Seis Sigma cuantificar la mejora de los procesos.

- **Descubra la relación entre mediciones del proceso (las x's) y los resultados del mismo (las y's).** De esta forma se pueden obtener dos resultados importantes, uno es mejorar la relación causa-efecto (cómo las variables de entrada afectan a las variables de salida). El otro resultado es la capacidad de predecir el futuro resultado del proceso al manipular las variables de entrada.

Tenga un proceso de medición

Seleccione el instrumento de medición adecuado

Use la frase de los maestros de obra “mida dos veces, corte una”.

Planee y mida el desempeño del proceso contra requerimientos del cliente.

- Seleccione lo que debe medirse.
- Desarrolle definiciones operacionales.
- Identifique las fuentes de datos.

Desarrolle medidas de posibles defectos e identifique oportunidades de mejora.

Datos Discretos vrs Datos Continuos

Datos continuos:

Mediciones que se pueden hacer en una escala continua divisible en un número infinito de sub-escalas. Todas las mediciones físicas para las que utilizamos un instrumento de medición con una escala definida.

Datos discretos:

Mediciones en unidades completas que no se pueden traslapar. También se le llama medición por atributos porque se cuentan los incidentes que tienen un atributo particular que está siendo estudiado.

- a. Utilice técnicas de muestreo.
- b. Aplique adecuados métodos para la recolección de datos asegurando precisión e integridad de datos

Datos Discretos vrs Datos Continuos

| | DISCRETOS (Contables) | | | | CONTINUOS |
|-------------|---|--|--|-------------------------------------|--------------------------------|
| | Binario | Categorías Desordenadas | Categorías Ordenadas | Cuenta | Medibles a una escala continua |
| Descripción | Clasificado en una u otra categoría | Sin Orden | Con Orden | Contados de Forma Discreta | |
| Ejemplo | Correcto / Incorrecto A tiempo / Tarde | Tipos de Productos (BMW - FIAT - AUDI - ETC) | Ranking de satisfacción al cliente de un call center | Número de errores en una aplicación | Tiempo - Temperatura - Peso |

Tomado de Mejora Continua de Iván Bulgarelli página 36.

Los Datos

Datos de Entrada

Datos del Proceso

Datos de Salida

Está relacionado con la calidad en la entrega del proveedor

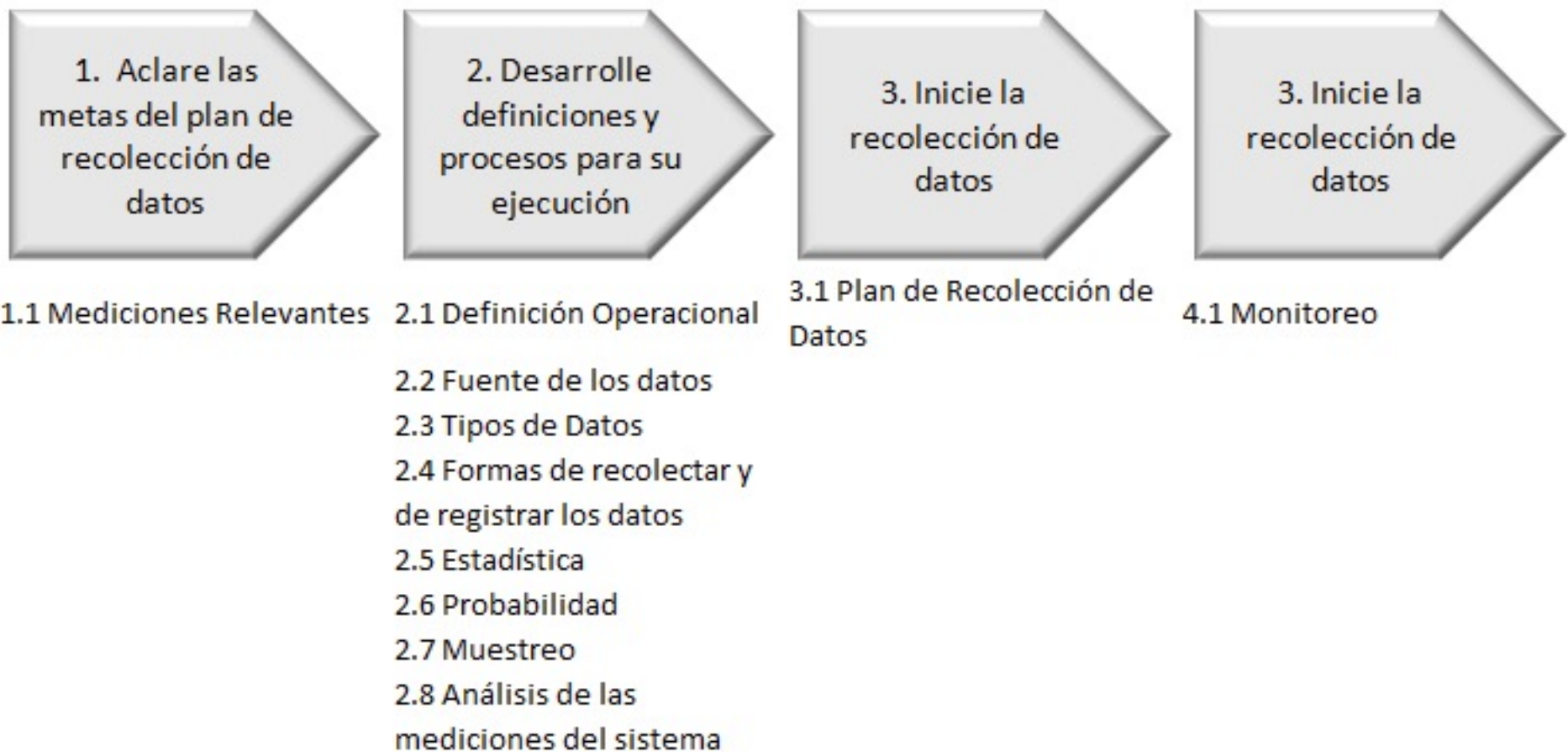
Los datos de procesos críticos dan el camino a seguir para encontrar los datos de salida deseados

Este tipo de dato nos mide si estamos alcanzando los CTC's y/o los CTB's

Pasos para la toma de los datos

1. Informar a colaboradores y superiores
2. Respetar las restricciones de la empresa
3. Entrenar a las personas encargadas
4. Contar con los recursos necesarios
5. Tener el Plan de Recolección de Datos
6. Recolecte los datos.

Plan de Recolección de Datos



Tomado de Mejora Continua de Iván Bulgarelli página 35.

Plan de Recolección de Datos

1. Que sea fácil y entendible
2. Personas diferentes recolectan de la misma forma
3. Se garantiza la consistencia
4. Fácil mantener el control de los datos



Plan de Recolección de Datos

¿Qué?

¿Cómo?

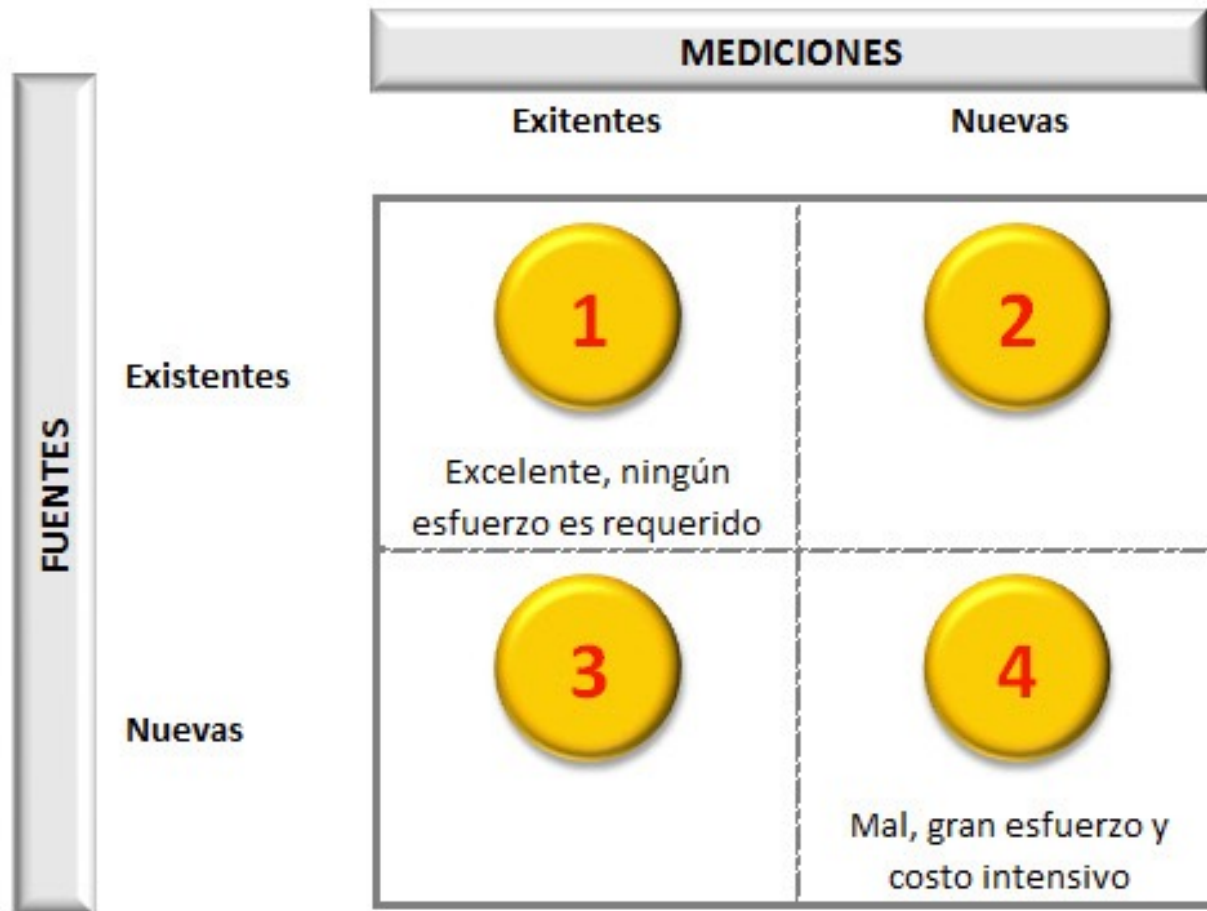
¿Quién?

¿Cuándo?

¿Dónde?

| Medición | Tipo de Medición (Salida/Entrada/Proceso) | Tipo de Dato (continuo/discreto) | Definición Operacional | Responsable | Fecha/Tiempo/Frecuencia | Fuente/Lugar |
|----------|--|-------------------------------------|---------------------------|-------------|-------------------------|--------------|
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

Fuente de los datos



Tomado de
Mejora
Continua de
Iván
Bulgarelli
página 36.

Escalas de Medición

| Escala | Descripción | Ejemplos |
|---------------|---|---|
| Nominal | Los números identifican y clasifican los objetos. | Número de marcas, tipos de tiendas, clasificación por sexo. |
| Ordinal | Los números indican las posiciones relativas de los objetos, pero no su magnitud o diferencias. | Posición en el mercad, clase social. |
| De Intervalos | Las diferencias entre los objetos pueden compararse. | Actitudes, opiniones, números en un índice. |
| De Relación | Las relaciones de los valores de la escala pueden calcularse. | Edad, ingreso, ventas, participación de mercado. |

Escalas de Medición

| Escala | Medida de Tendencia central | Dispersión | Prueba de Significancia |
|---------------|-----------------------------------|-------------------------|---|
| Nominal | Moda | Solamente información | Chi cuadrado |
| Ordinal | Mediana | Porcentajes | Prueba de Signo o Prueba de Corrida |
| De Intervalos | Media Aritmética | Desviación Estándar | Prueba t Prueba F Análisis de Correlación |
| De Relación | Media Geométrica o Media Armónica | Porcentaje de Variación | -- |

Muestreo

El muestreo es el proceso mediante el cual se hace la recolección de los datos.

Una muestra es una sección de la población que si es representativa, permite hacer conclusiones acerca de la población.

Proceso de Muestreo

| Situación | Propósito | Consideraciones | Método |
|-----------|--|--|--|
| Proceso | Se mide para tomar acciones o proyectar: -Hay estabilidad? -El proceso es capaz? -Se puede mejorar? | Exactamente que parte del proceso se va a muestrear? Con qué frecuencia? En cuáles grupos? Se tiene representatividad? Cuál es el costo? | Muestreo de Grupos Muestrear Sistemáticamente |
| Población | Describir, cuantificar, analizar características | Precisión Grado de variación del proceso Nivel de confiabilidad Muestra representativa Cuál es el costo? | Muestreo Aleatorio Muestreo Aleatorio Estratificado Muestrear Sistemáticamente |

Mínimos

| Información Objetivo | Tamaño Mínimo de la Muestra |
|----------------------|------------------------------------|
| Promedio | 30 - 40 |
| Desviación Estándar | 30 - 40 |
| Porción defectuosa | 100, $np \geq 5$ & $n(1-p) \geq 5$ |
| Diagrama de Pareto | 50 |
| Regresión | 30 |
| Gráficos de Control | 30 |



Estadística Descriptiva

Capacidad de Procesos I

Baje el archivo de excel con el nombre Sesión 05 medir y el formulario del curso.

Estadística Descriptiva

1. Describe un conjunto de datos.
2. Elaboración de cuadros y gráficos.
3. Cálculo de medidas de resumen.



Cuadros Estadísticos

Cuadro 9

Costa Rica: Distribución absoluta del abstencionismo en la elección presidencial de 1998 por provincias según sexo

| Provincia | TOTAL | Porcentaje | Hombres ^{a)} | Mujeres ^{b)} |
|--------------|----------------|--------------|-----------------------|-----------------------|
| San José | 241.628 | 39,3 | 125.532 | 116.096 |
| Alajuela | 98.495 | 16,0 | 53.188 | 45.307 |
| Puntarenas | 64.412 | 10,5 | 34.725 | 29.687 |
| Limón | 61.271 | 10,0 | 33.899 | 27.372 |
| Cartago | 59.173 | 9,6 | 31.435 | 27.738 |
| Heredia | 52.141 | 8,5 | 28.351 | 23.790 |
| Guanacaste | 36.947 | 6,0 | 20.334 | 16.613 |
| TOTAL | 614.067 | 100,0 | 327.464 | 286.603 |

a) Representa el 25% del padrón electoral de hombres

b) Representa el 32% del padrón electoral de mujeres

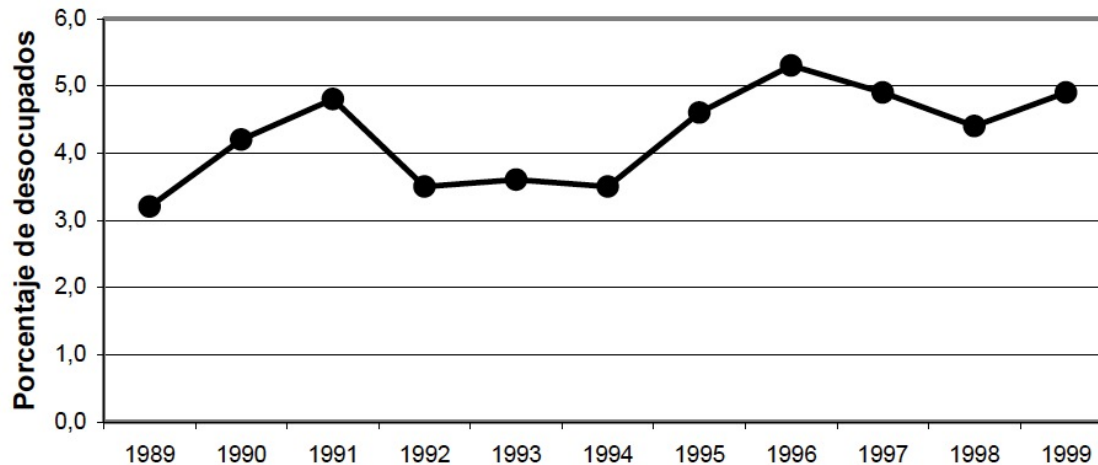
Fuente: Tribunal Supremo de Elecciones. *Estadísticas del Sufragio 1998*. Imprenta Nacional. San José, Costa Rica, 2001

Permite
mostrar gran
cantidad de
datos

Gráficos Lineales

Gráfico 7

Costa Rica: Porcentaje de desocupados (tasa de desempleo abierto) por sexo, 1989-1999

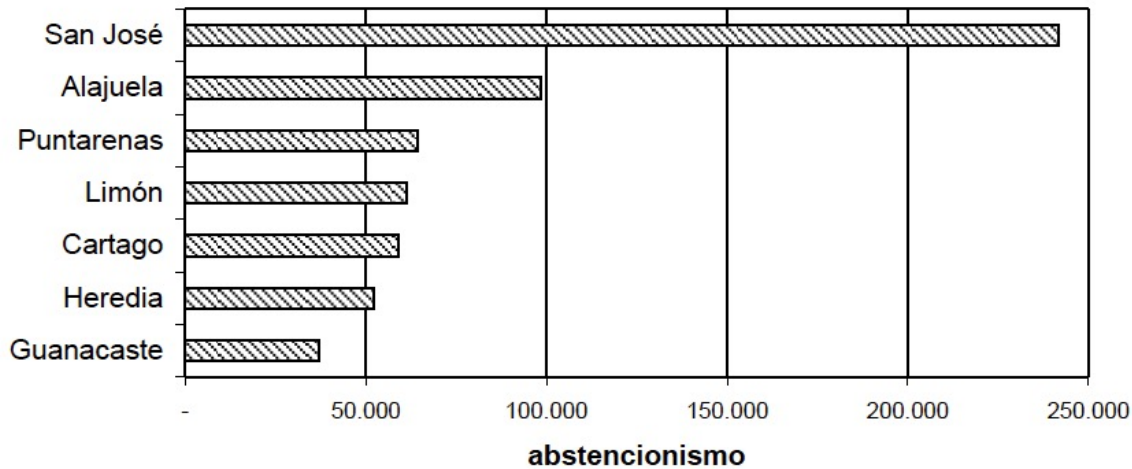


Se utilizan para comparar el comportamiento de una variable en el tiempo

Fuente: Elaboración propia con base en datos del Sexto Informe del Estado de la Nación. San José, Costa Rica, 1999

Gráficos de Barras Horizontales

Gráfico 8
Costa Rica: Distribución del total de abstencionistas en la elección presidencial de 1998, según provincias

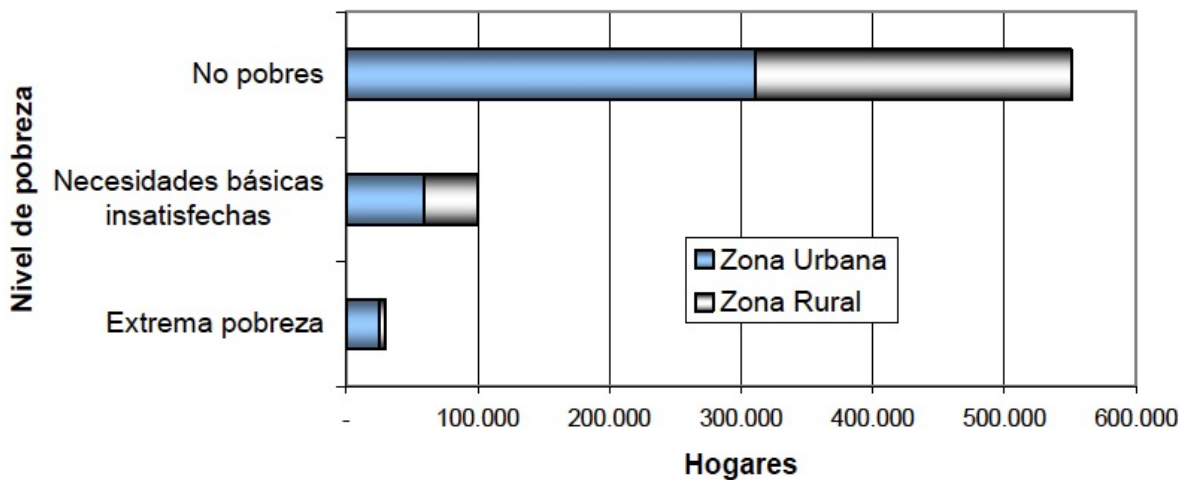


Muy útiles cuando los nombres de las categorías son muy extensos

Fuente: Elaboración propia con base en datos del Tribunal Supremo de Elecciones. *Estadísticas del sufragio 1998*. Imprenta Nacional. San José, Costa Rica, 2001

Gráficos de Barras Compuestas

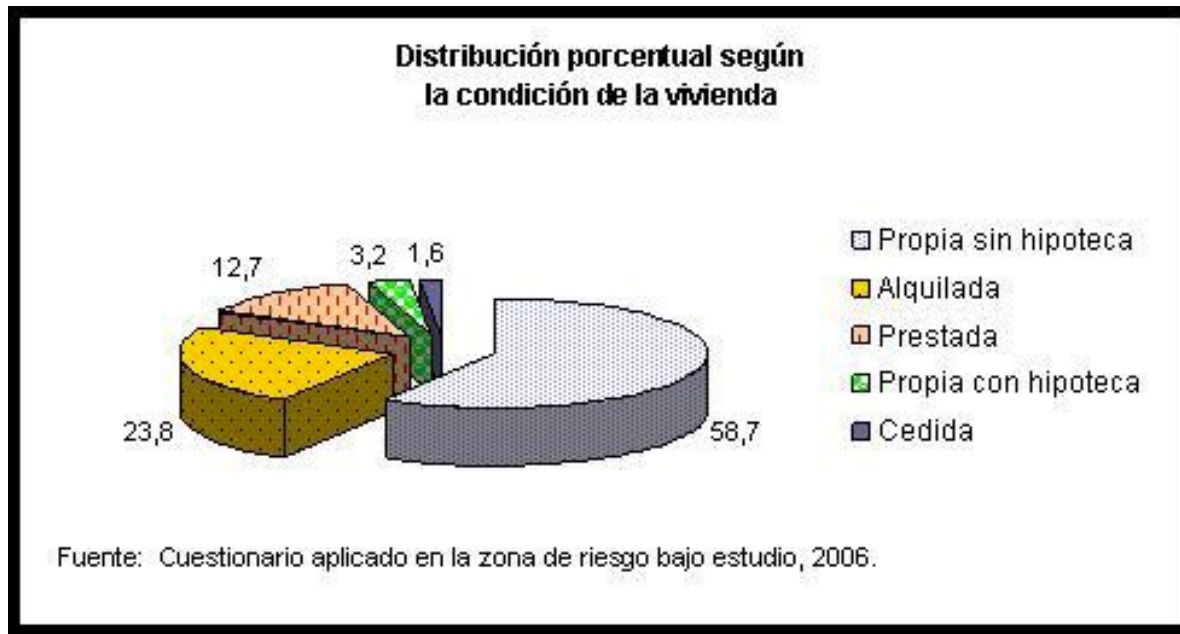
Gráfico 7
Costa Rica: Número de hogares según niveles de pobreza por zona, 1998



Permite combinar dos categorías en un mismo gráfico.

Fuente: Elaboración propia con base en datos del Sexto Informe del Estado de la Nación. San José, Costa Rica, 1999

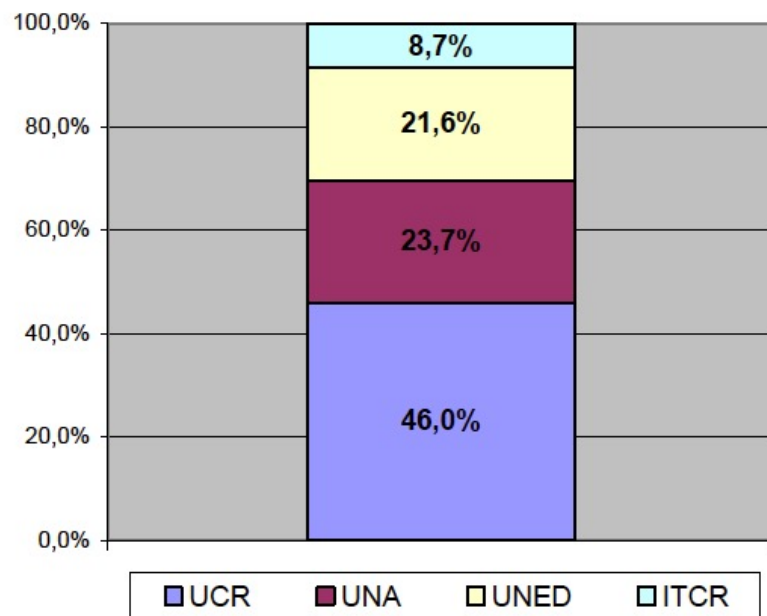
Gráfico Circular



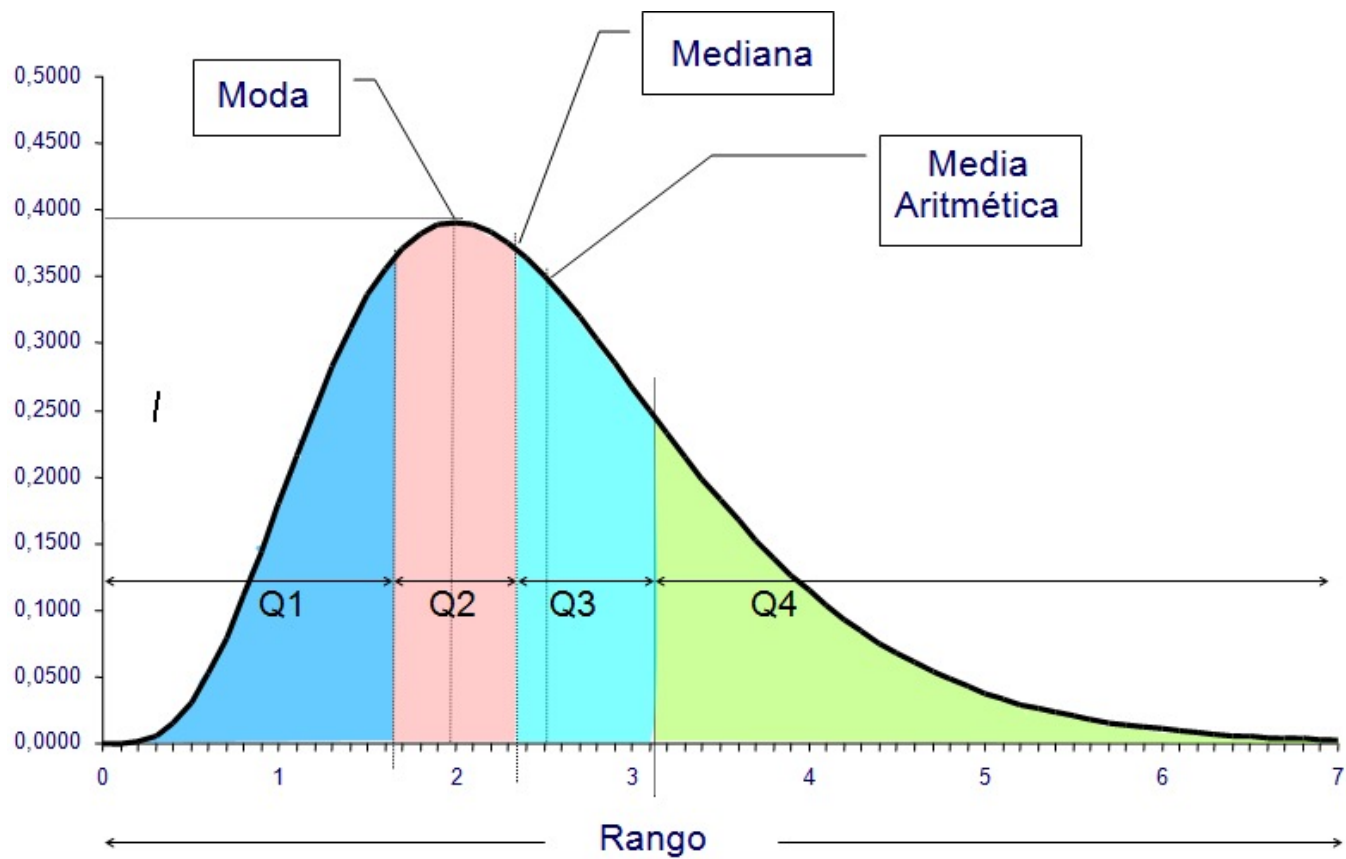
Permiten mostrar una cantidad limitada de categorías.
No más de 6.

Gráfico de Barra 100%

Gráfico 7
Costa Rica: Porcentaje de diplomas otorgados por Instituciones Públicas de Educación Superior, 1998



Permiten mostrar una cantidad limitada de categorías.
No más de 5.



Fuente: Allende H y Ahumada S, ILI-280

MEDIDAS DE POSICIÓN

La Moda

Corresponde al valor observado con mayor frecuencia:

Si lo datos fueran:

Ana, Pedro, Ana, Luis, Pedro, Ana

333, 33, 3, 21, 12, 121

Alto, Bajo, Alto, Mediano, Bajo

11, 325, 12, 45, 1, 14, 1

La MODA sería:

La Moda para Datos Agrupados

Se debe determinar la clase modal, es decir, la clase cuya frecuencia relativa es mayor a la de las demás.

$$Mo = Li + \frac{d_1}{d_1 + d_2} \cdot c$$

En donde:

Li es el límite inferior de la clase modal

d_1 es la diferencia entre la frecuencia de la clase modal y la anterior

d_2 es la diferencia entre la frecuencia de la clase modal y la posterior

C es el intervalo de clase modal

La Mediana

$$P_{Me} = \frac{n+1}{2}$$

*donde P_{Me} es la posición de la mediana
y n es el número de datos*

Corresponde al valor que se localiza en el centro de los datos ordenados:

Si lo datos fueran:

Ana, Pedro, Ana, Luis, Pedro, Ana

333, 33, 3, 21, 12, 121

Alto, Bajo, Alto, Mediano, Bajo

11, 325, 12, 45, 1, 14, 1

La MEDIANA sería:

La Mediana para Datos Agrupados

Se debe determinar la clase de mediana, es decir la clase cuya frecuencia acumulada incluye a la mediana de los datos.

$$Me = Li + \frac{\frac{n}{2} - Fa}{f_i} \cdot c$$

En donde:

Li es el límite inferior de la clase mediana

n número de datos

Fa frecuencia acumulada de la clase anterior

f_i frecuencia absoluta donde está la mediana

c intervalo de clase

La Media Aritmética

También llamada promedio o media, de un conjunto infinito de números es el valor característico de una serie de datos cuantitativos, objeto de estudio que parte del principio de la esperanza matemática o valor esperado, se obtiene a partir de la suma de todos sus valores dividida entre el número de sumandos. Cuando el conjunto es una muestra aleatoria recibe el nombre de media muestral siendo uno de los principales estadísticos muestrales.

$$\bar{X} = \frac{5+6+7+4+9}{5}$$

La Media Aritmética

Con Datos NO Agrupados

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

x_i es el dato i

n es el número de datos

La Media Aritmética

Con Datos Agrupados

$$\overline{X} = A + \frac{(\sum f_o \times d) \times i}{n}$$

- En donde:
- d = desviación del punto medio con respecto a la posición de la media supuesta, es medida en unidades de intervalo de clase
- i = amplitud o intervalo de clase
- A = punto medio de la clase que contiene la media supuesta (clase de $d=0$)
- f_o = Frecuencias observadas en número de clases
- n = tamaño de la muestra

Ejercicio

Calcule la moda, la mediana y la media para estos datos.

| N° clase | LI | LS | Frecuencia |
|----------|------|-------|------------|
| 1 | 49.5 | 56.5 | 7 |
| 2 | 56.5 | 63.5 | 4 |
| 3 | 63.5 | 70.5 | 4 |
| 4 | 70.5 | 77.5 | 7 |
| 5 | 77.5 | 84.5 | 11 |
| 6 | 84.5 | 91.5 | 9 |
| 7 | 91.5 | 98.5 | 7 |
| 8 | 98.5 | 105.5 | 1 |

La clase modal corresponde a la clase comprendida entre 77,5 y 84,5 pues es la clase que posee la mayor frecuencia entre todas las clases, igual a 11.

$$L_i = 77,5$$

$$d_1 = 11 - 7 = 4$$

$$d_2 = 11 - 9 = 2$$

$$c = 84,5 - 77,5 = 7$$

$$Mo = L_i + \frac{d_1}{d_1 + d_2} * c = 77,5 + \frac{4}{4 + 2} * 7 = 82,17 \approx 82$$

Para poder encontrar la clase mediana, debemos calcular la posición de la mediana.

$$P_{Me} = \frac{n+1}{2} =$$

Para poder encontrar la clase mediana, debemos calcular la posición de la mediana.

$$P_{Me} = \frac{n+1}{2} = \frac{50+1}{2} = 25,5$$

La clase mediana corresponde a la clase que contenga los valores que se ubican entre la posición 25 y la posición 26.

Para poder encontrar dicha posición debemos calcular la frecuencia acumulada de cada clase.

| N de Clase | Li | Ls | Frecuencia | Frecuencia Acumulada |
|------------|------|-------|------------|----------------------|
| 1 | 49.5 | 56.5 | 7 | 7 |
| 2 | 56.5 | 63.5 | 4 | 11 |
| 3 | 63.5 | 70.5 | 4 | 15 |
| 4 | 70.5 | 77.5 | 7 | 22 |
| 5 | 77.5 | 84.5 | 11 | 33 |
| 6 | 84.5 | 91.5 | 9 | 42 |
| 7 | 91.5 | 98.5 | 7 | 49 |
| 8 | 98.5 | 105.5 | 1 | 50 |

$$Me = L_i + \frac{n/2 - F_a}{f_i} \cdot c$$

$$L_i = 77,5$$

$$n = 50$$

$$F_a = 22$$

$$f_i = 11$$

$$c = 56,5 - 49,5 = 7$$

$$Me = 77,5 + \frac{50/2 - 22}{11} \cdot 7 = 79,41 \approx 79,4$$

Para calcular la media es necesario calcular el punto medio de cada intervalo de clase

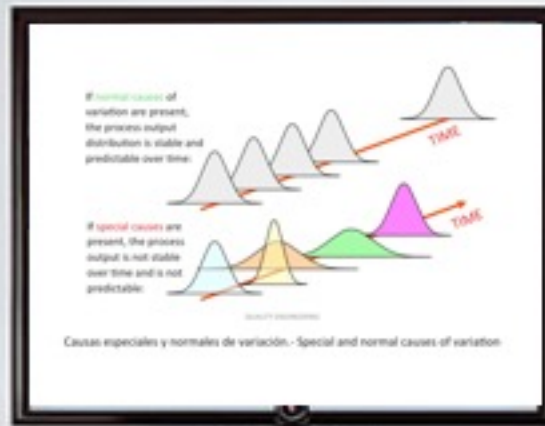
| N de Clase | Li | Ls | Frecuencia | Frecuencia Acumulada | Punto Medio | d | fo x d | fo x d^2 |
|------------|------|-------|------------|----------------------|-------------|---|--------|----------|
| 1 | 49.5 | 56.5 | 7 | 7 | 53.00 | | | |
| 2 | 56.5 | 63.5 | 4 | 11 | 60.00 | | | |
| 3 | 63.5 | 70.5 | 4 | 15 | 67.00 | | | |
| 4 | 70.5 | 77.5 | 7 | 22 | 74.00 | | | |
| 5 | 77.5 | 84.5 | 11 | 33 | 81.00 | | | |
| 6 | 84.5 | 91.5 | 9 | 42 | 88.00 | | | |
| 7 | 91.5 | 98.5 | 7 | 49 | 95.00 | | | |
| 8 | 98.5 | 105.5 | 1 | 50 | 102.00 | | | |

$$\bar{X} = A + \frac{(\sum fo \times d) \times i}{n}$$

Para calcular la media es necesario calcular el punto medio de cada intervalo de clase

| N de Clase | Li | Ls | Frecuencia | Frecuencia Acumulada | Punto Medio | d | fo x d | fo x d^2 |
|------------|------|-------|------------|----------------------|-------------|----|--------|----------|
| 1 | 49.5 | 56.5 | 7 | 7 | 53.00 | -4 | -28 | 112 |
| 2 | 56.5 | 63.5 | 4 | 11 | 60.00 | -3 | -12 | 36 |
| 3 | 63.5 | 70.5 | 4 | 15 | 67.00 | -2 | -8 | 16 |
| 4 | 70.5 | 77.5 | 7 | 22 | 74.00 | -1 | -7 | 7 |
| 5 | 77.5 | 84.5 | 11 | 33 | 81.00 | 0 | 0 | 0 |
| 6 | 84.5 | 91.5 | 9 | 42 | 88.00 | 1 | 9 | 9 |
| 7 | 91.5 | 98.5 | 7 | 49 | 95.00 | 2 | 14 | 28 |
| 8 | 98.5 | 105.5 | 1 | 50 | 102.00 | 3 | 3 | 9 |
| | | | | | | | -29 | 217 |

$$\bar{X} = A + \frac{(\sum fo \times d) \times i}{n} \quad 81 + \frac{(-29 \times 7)}{50} = 76,94$$



MEDIDAS DE VARIABILIDAD

Para Datos No Agrupados

Rango

$$\text{Max} - \text{Min}$$

Desviación Estándar

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

Varianza

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

Coeficiente de Variación

$$CV = \frac{s}{\bar{x}} \cdot 100$$



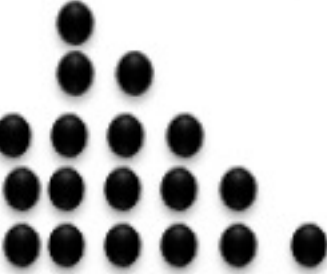

Para Datos Agrupados


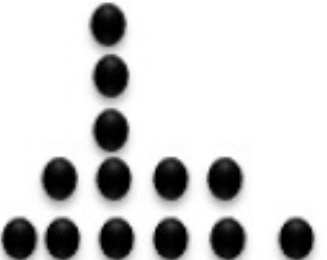
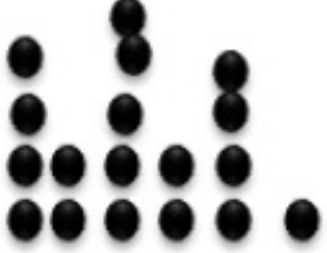
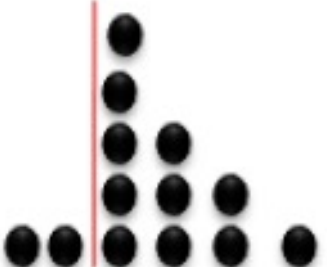
$$S = i \sqrt{\frac{\sum fo \times d^2}{n} - \frac{(\sum fo \times d)^2}{n^2}}$$

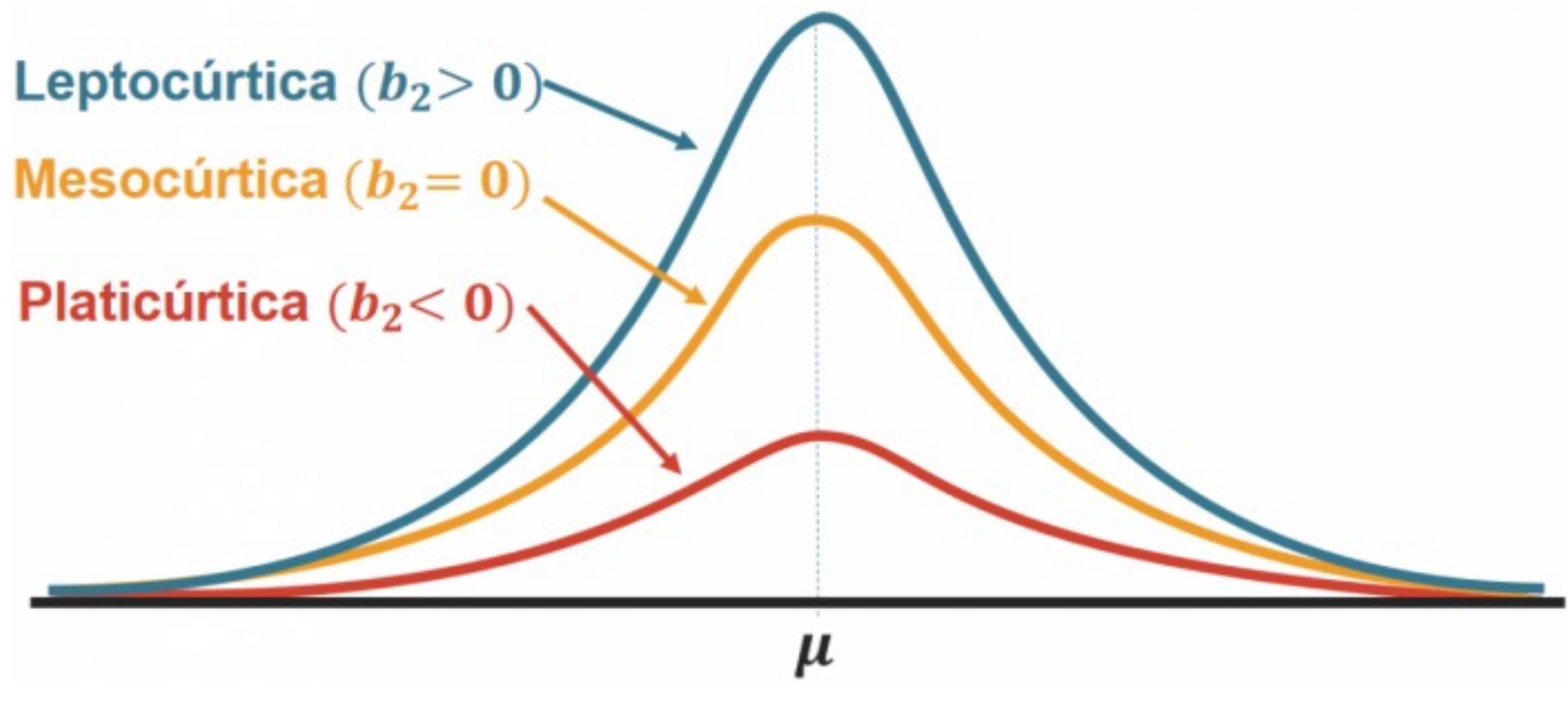
Límites reales o naturales

Límites reales

Se obtienen con $\mu - 3\sigma$ y $\mu + 3\sigma$, e indican de dónde a dónde varía la salida de un proceso.

| Histograma | Descripción | Interpretación | Cómo Manejarlo |
|---|-------------------------------------|---|--|
|  <p>Diagrama de histograma simétrico en forma de campana. Las barras están representadas por círculos negros. Hay 1 círculo en la primera posición, 2 en la segunda, 3 en la tercera, 4 en la cuarta, 3 en la quinta, 2 en la sexta y 1 en la séptima. Esto forma una curva simétrica y equilibrada.</p> | Curva simétrica en forma de campana | No existen causas especiales. Los datos pueden venir de un proceso estable | |
|  <p>Diagrama de histograma bimodal con dos picos. Hay 1 círculo en la primera posición, 2 en la segunda, 3 en la tercera, 1 en la cuarta, 2 en la quinta, 3 en la sexta y 4 en la séptima. Esto muestra dos grupos de datos separados.</p> | Bimodal, dos picos | Los datos pueden provenir de poblaciones mezcladas | Examinar antes de realizar el análisis |
|  <p>Diagrama de histograma asimétrico. Hay 1 círculo en la primera posición, 2 en la segunda, 3 en la tercera, 4 en la cuarta, 3 en la quinta, 2 en la sexta y 1 en la séptima. La curva se inclina hacia la derecha, indicando un valor atípico o una cola larga.</p> | Curva asimétrica | Los datos podrían no ser normales, por lo que no se pueden hacer suposiciones sencillas | Considere transformación de datos |
|  <p>Diagrama de histograma con poca dispersión y pocos datos. Hay 1 círculo en la primera posición, 2 en la segunda, 3 en la tercera y 4 en la cuarta. El grupo de datos es muy estrecho y tiene una muestra pequeña.</p> | Poca dispersión. Pocos Datos | El instrumento de medición no es lo suficientemente sensible | Mejore el proceso de medición |

| | | | |
|---|--|--|---|
|  | <p>Especificación del instrumento. Gran cantidad de información en el valor mínimo o máximo</p> | <p>El instrumento de medición no es capaz de leer el espectro total de los datos</p> | <p>Extender el rango de frecuencia del instrumento</p> |
|  | <p>Gran cantidad de datos en un rango específico</p> | <p>El instrumento puede estar dañado</p> | <p>Observe las razones físicas que provocan este estado</p> |
|  | <p>Patrón de diente. Quién ejecuta favorece valores específicos</p> | <p>Curvas Inconsistentes</p> | <p>Estandarizar el proceso escrito</p> |
|  | <p>Mayoría de datos agrupados en un límite de especificación, ningún dato o pocos luego del límite</p> | <p>Es posible que los datos hayan sido manipulados o inventados</p> | <p>Entrenar en el proceso de la toma de datos</p> |

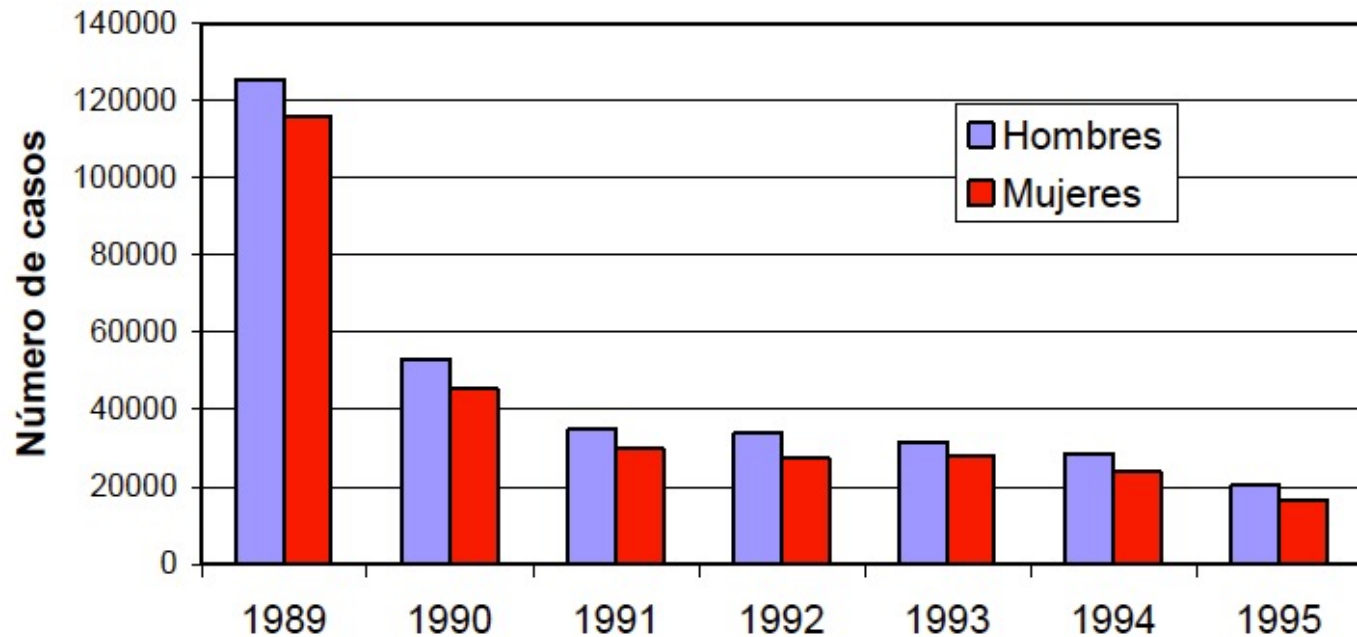


MEDIDAS DE FORMA

Forma de la Distribución

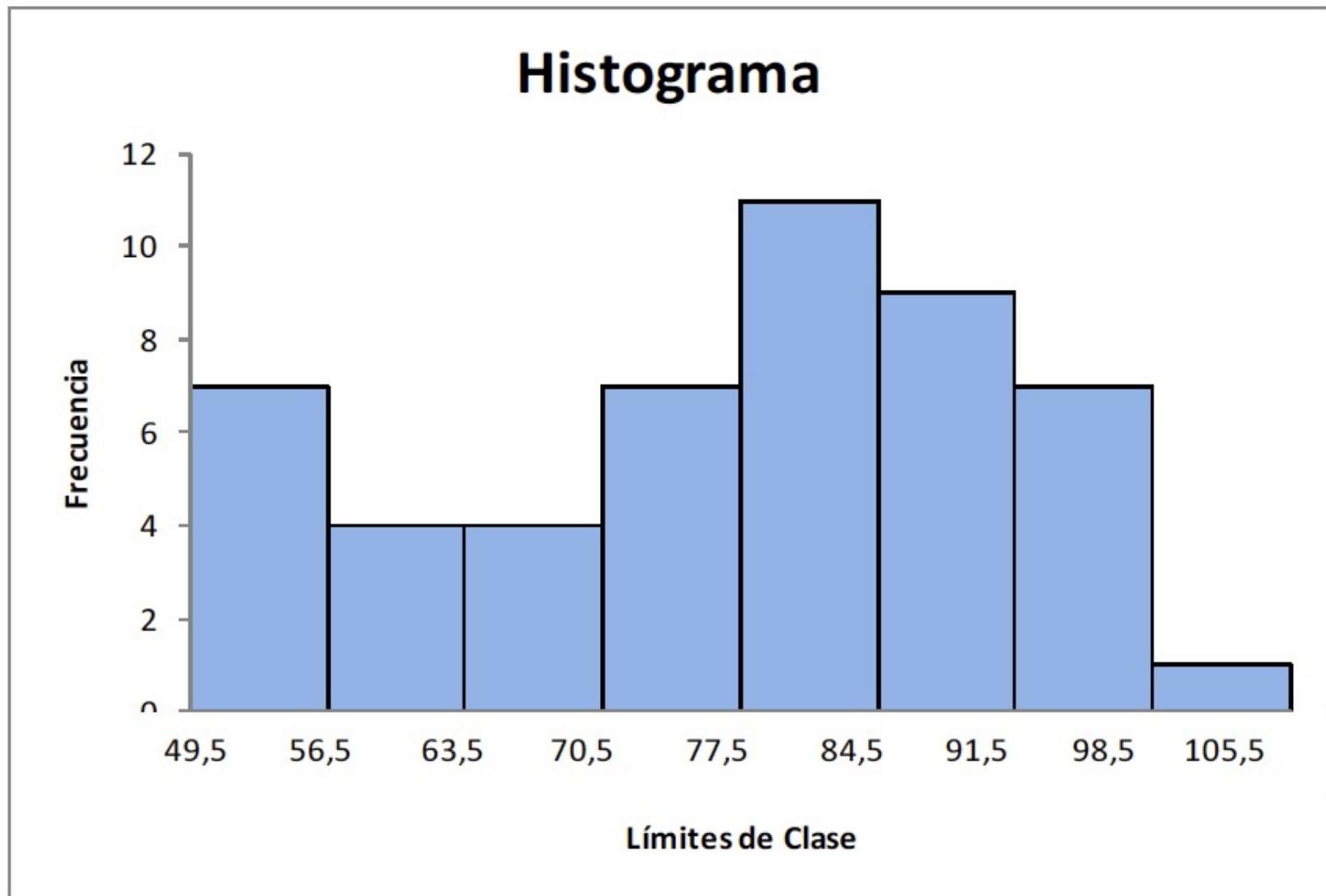
1. Se obtiene agrupando los datos en categorías o clases en una distribución de frecuencias.
2. Corresponden a la forma en que los datos se comportan.
3. Permite visualizar en forma conjunta los valores centrales y la dispersión de los datos.
4. En variables cualitativas se emplean gráficos de barras
5. En variables cuantitativas se utiliza el histograma

Gráfico 7
Costa Rica: Cantidad de desempleados, 1989-1995



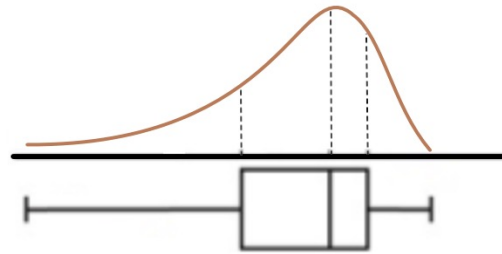
Fuente: Elaboración propia con base en datos del Sexto Informe del Estado de la Nación. San José, Costa Rica, 1999

El gráfico de barras se emplea para variables nominales y ordinales. En el caso de variables nominales, los valores deben ordenarse de mayor a menor frecuencia, mientras que para variables ordinales se debe respetar el orden establecido por la variable.

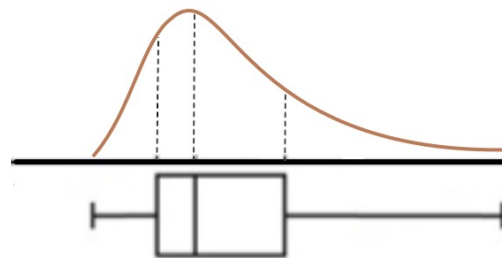


El Histograma se emplea para variables continuas. Se representa con barras sin separación entre sí. Las barras se denominan «clases», cuyo ancho se determina por el «intervalo de clase». Entre una clase y otra se establece un «límite de clase» el cual contiene un decimal más a los datos originales para evitar dudas cuando un dato quede en medio de dos clases.

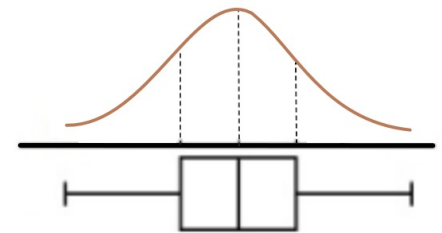
Sesgo



Distribucion Asimetrica Negativa



Distribucion Asimetrica Positiva



Distribucion Simetrica

Sesgo

Es una medida numérica de la asimetría en la distribución de un conjunto de datos.

$$\text{Sesgo} = \frac{n \sum_{i=1}^n (x_i - \bar{X})^3}{(n-1)(n-2)S^3}$$

$$\text{Sesgo estandarizado} = \frac{\text{Sesgo}}{\sqrt{\frac{6}{n}}}$$

Sesgo para datos agrupados

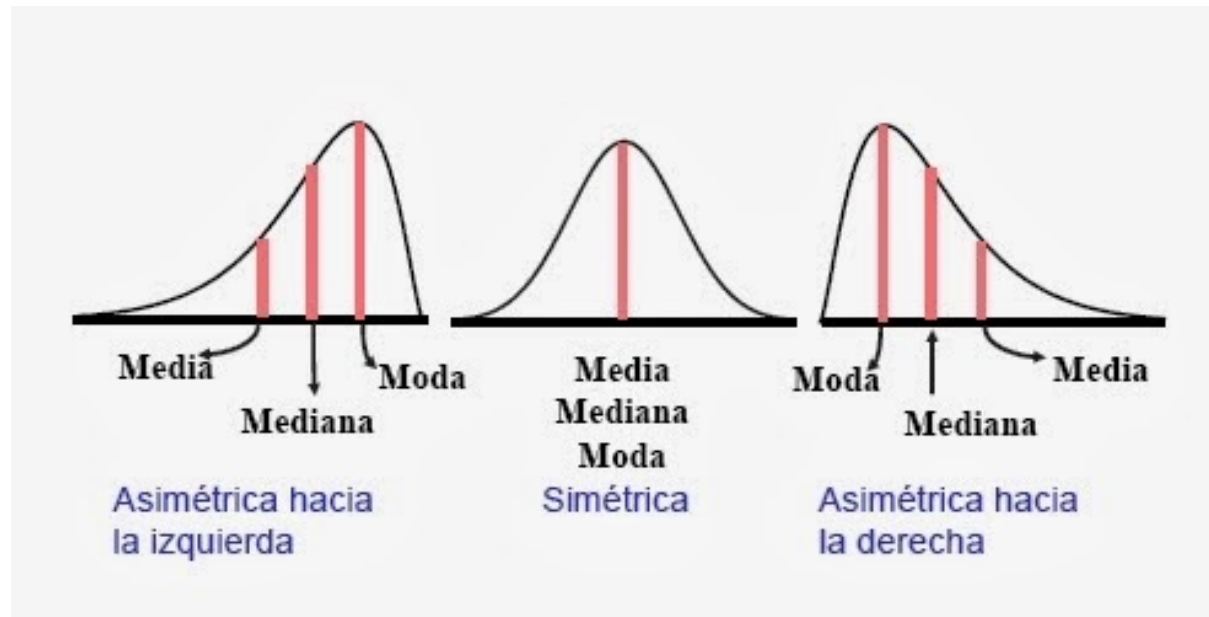
$$Sesgo = \frac{1}{n} * \frac{\sum (M_i - \bar{x})^3 * F_o}{S^3}$$

Donde: M_i es el punto medio de cada límite de clase
 S es la desviación estándar
 F_o la frecuencia observada
 n el tamaño de la muestra
 \bar{x} barra la media de la muestra

Sesgo

Para los datos que siguen una distribución normal, el valor del sesgo estandarizado debe de caer dentro de $(-2,2)$. Si el tamaño de la muestra supera 100 y el sesgo estandarizado está fuera de este intervalo es evidencia que la distribución de los datos tiene un sesgo y NO es normal.

El signo del sesgo indica el lado para el que la cola es más larga. Hacia la izquierda el signo sería negativo y hacia la derecha el signo sería positivo.



-

+

Curtosis

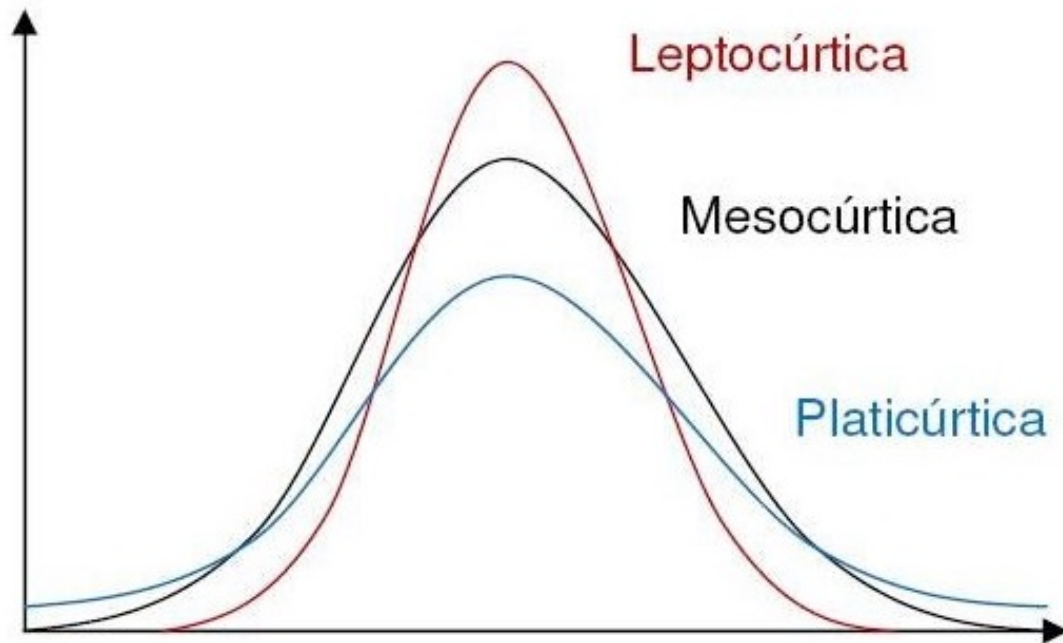
Mide que tan elevada o plana es la curva de la distribución de unos datos respecto a la distribución normal.

$$\text{Curtosis} = \frac{n(n+1) \sum_{i=1}^n (x_i - \bar{x})^4}{(n-1)(n-2)(n-3)S^4} - \frac{3(n-1)}{(n-2)(n-3)}$$

$$\text{Curtosis estandarizado} = \frac{\text{curtosis}}{\sqrt{\frac{24}{n}}}$$

Curtosis

La curtosis estandarizada debe estar dentro de $(-2,2)$. Si el tamaño de la muestra es mayor a 100 y el estadístico cae fuera de este intervalo, será evidencia de que la distribución de los datos NO es normal.



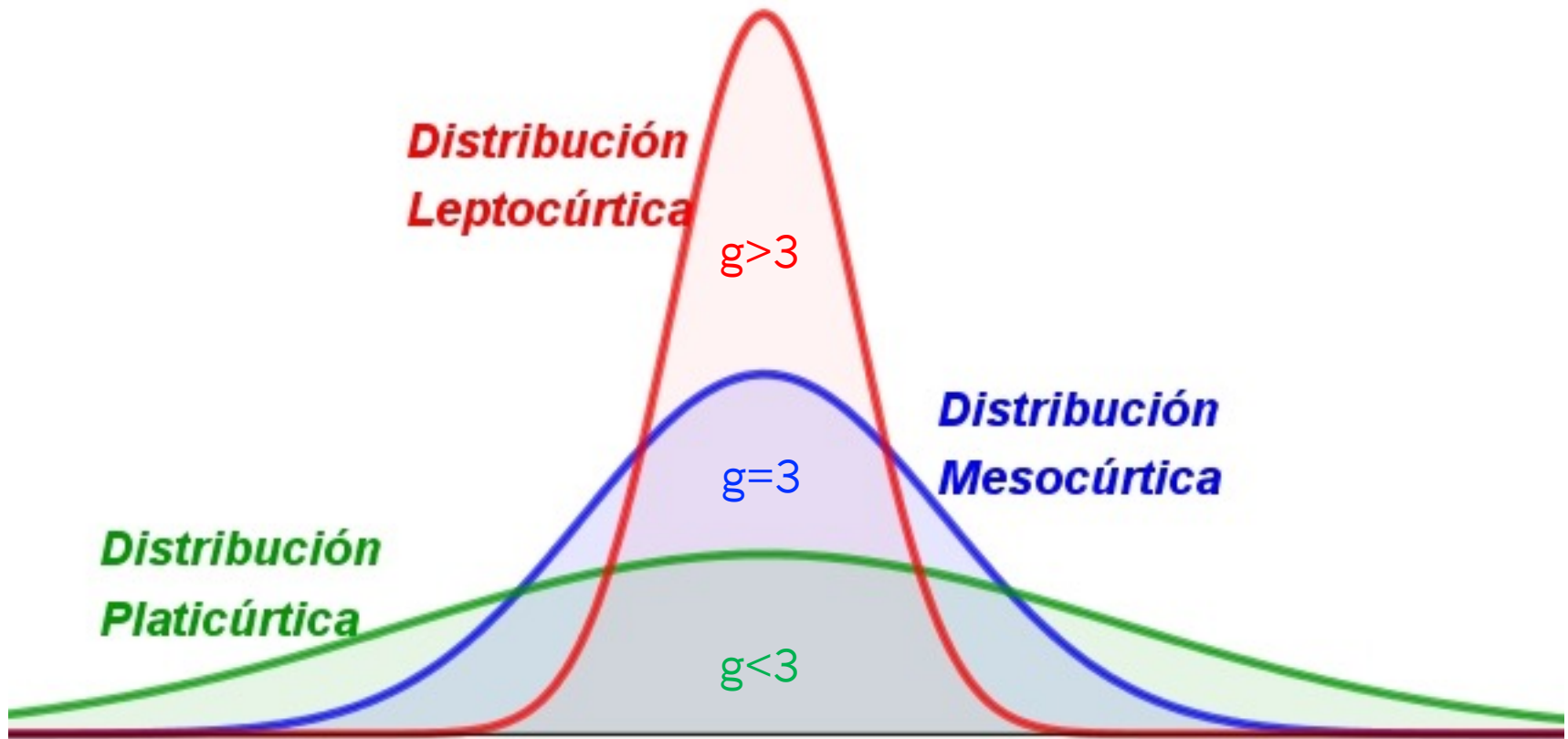
- Si los datos están muy concentrado hacia la media, la distribución es leptocúrtica (curtosis mayor a 0).
- Si los datos están muy dispersos, la distribución es platicúrtica (curtosis menor a 0).
- El comportamiento normal exige que la curtosis sea igual a 0 (distribución mesocúrtica).

Curtosis para datos agrupados

$$\text{Curtosis} = \frac{\frac{1}{n} \sum (M_i - \bar{x})^4 * F_o}{S^4}$$

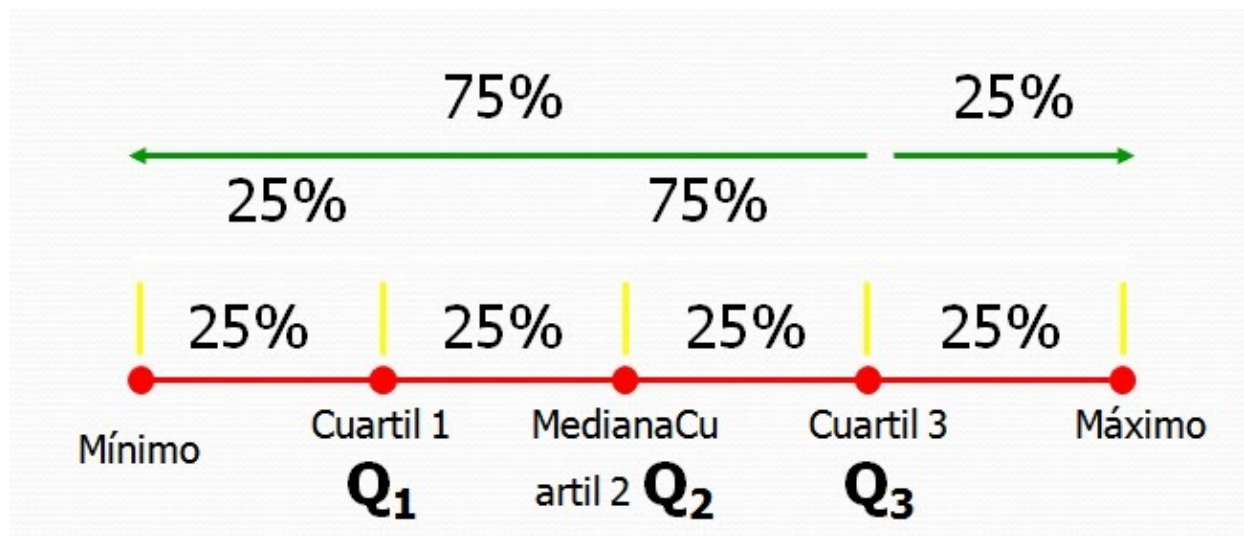
Donde: M_i es el punto medio de cada límite de clase
 S es la desviación estándar
 F_o la frecuencia observada
 n el tamaño de la muestra
 \bar{x} la media de la muestra

Curtosis para datos agrupados



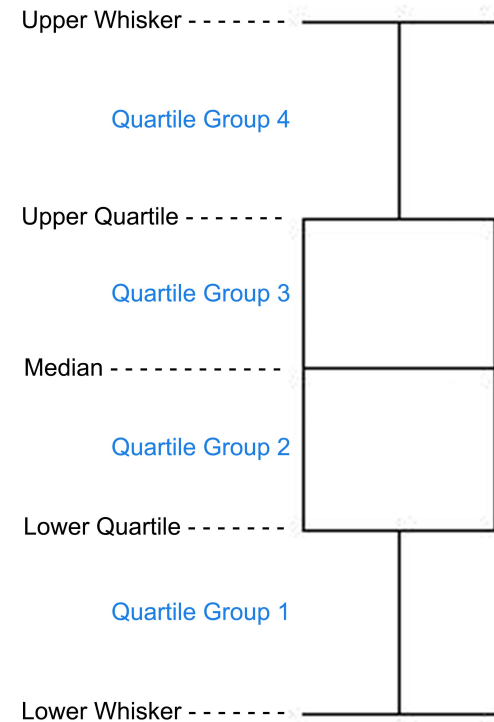
Cuantiles (percentiles)

Medidas de localización que separan por magnitud un conjunto de datos en cierto número de grupos o partes que contienen la misma cantidad de datos, por ejemplo los deciles dividen los datos en grupos de 10.

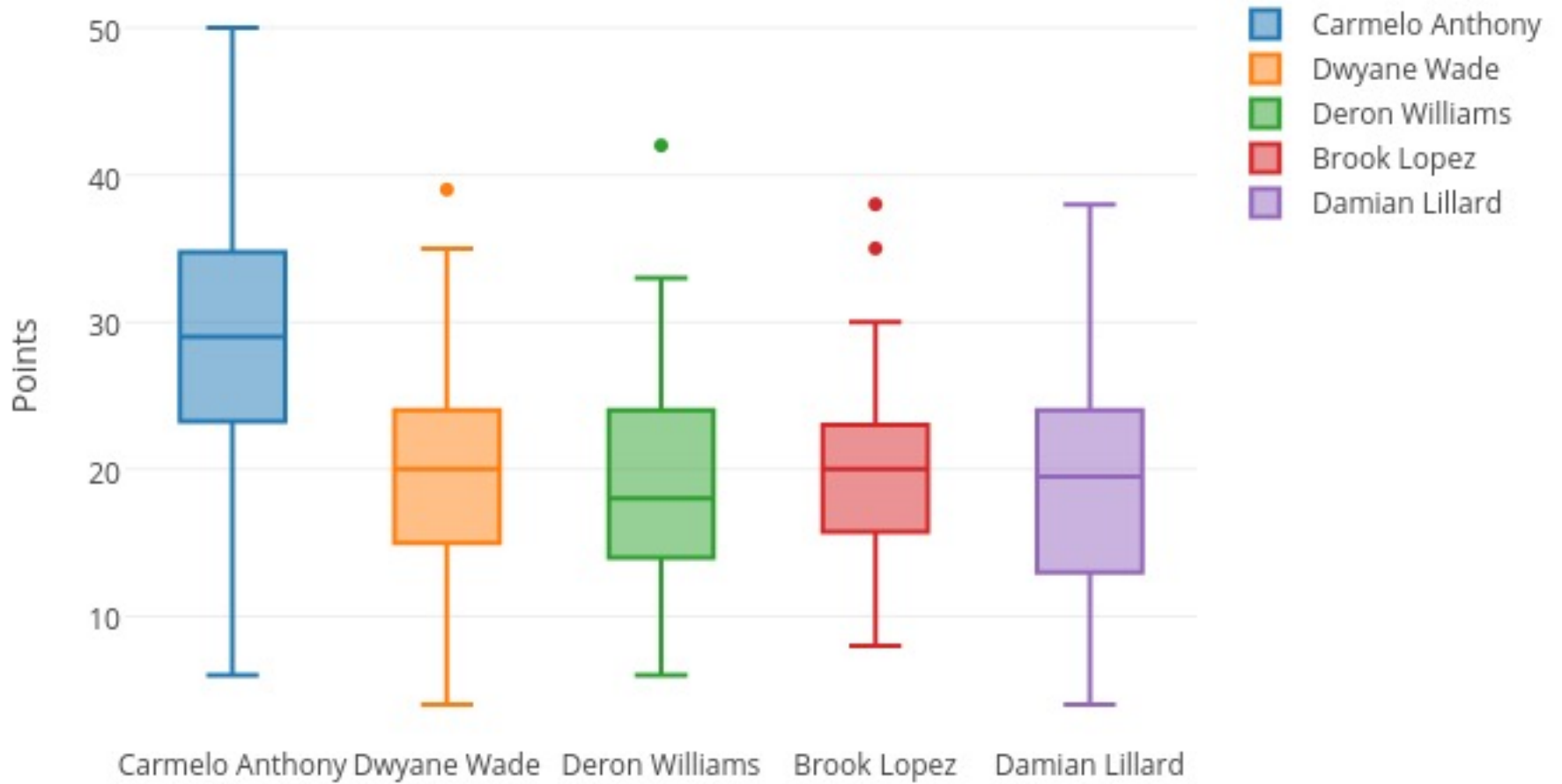


Diagramas de Caja o Box Plot

Permite ver fácilmente características de un grupo de datos, como la variación (largo de la caja y los bigotes), donde está el valor mínimo, máximo, cuartiles, etc. Permite comparar exactitud y precisión de diferentes elementos con alguna relación.



Points Scored Per NBA Game



Pregunta número 1(15 pts.) Los siguientes datos representan las mediciones de viscosidad de los últimos tres meses de un producto lácteo. El objetivo es tener una viscosidad de 80 ± 10 cps.

| | Producto lácteo | | | | | | | |
|----|-----------------|----|----|----|----|----|----|--|
| 84 | 86 | 82 | 86 | 78 | 82 | 84 | 80 | |
| 86 | 82 | 82 | 90 | 84 | 81 | 82 | 78 | |
| 78 | 80 | 83 | 81 | 86 | 86 | 83 | 84 | |
| 79 | 86 | 87 | 83 | 81 | 84 | 83 | 83 | |
| 81 | 80 | 78 | 87 | 85 | 82 | 86 | 83 | |
| 85 | 84 | 84 | 78 | 81 | 79 | 84 | 84 | |
| 83 | 80 | 82 | 82 | 82 | 82 | 82 | 82 | |
| 81 | 83 | 88 | 81 | 84 | 87 | 82 | 81 | |
| 77 | 86 | 83 | 83 | 82 | 80 | 82 | 81 | |
| 79 | 85 | 82 | 84 | 84 | 80 | 81 | 76 | |

- Obtenga las medidas de tendencia central y con base en ellas señale si la tendencia central del proceso es adecuada
- Calcule la desviación estándar y una aproximación de los límites reales. A partir de éstos decida si la variabilidad de los datos es aceptable.
- Construya una gráfica de capacidad de este proceso (histograma con tolerancias) y dé una primera opinión sobre la capacidad.
- Utilizando el sesgo y curtosis estandarizadas y la evidencia gráfica, ¿qué puede decir respecto a la normalidad de los datos?