**PRESENTACION**



El presente material ofrece los lineamientos generales presentados en las sesiones: " Estadísticas Exploratorias " y " Estimación Econométrica " durante el " CURSO SOBRE EL PAPEL DE LA SOCIOECONOMIA EN LA GENERACION DE TECNOLOGIA AGRICOLA ", realizado en CIAT, entre Septiembre 16 y Octubre 11 de 1991.

Se pretende ofrecer a los usuarios una copia de los acetatos que se proyectan, con el fin de que puedan registrar en ella explicaciones adicionales, comentarios, ejemplos o cualquier elemento que contribuya a la claridad del tema considerado. De esta manera, al ser un material para uso paralelo a una explicación, permite a quien vuelva sobre sus páginas recordar un concepto o reconsiderar un ejemplo en forma directa.

Debido a la diferencia entre participantes en téminos de conocimientos estadísticos, este manual debe plantear un bosquejo del tema, permitiendo detenerse o avanzar rápidamente según el grupo lo requiera. Por estas razones, no recomendamos su uso para alguien que no haya estado presente en las exposiciones. Los enunciados de los conceptos son muy generales y al no existir interacción con el expositor, no hay la posibilidad de hacer correciones, ampliaciones, o de ofrecer mayor claridad sobre puntos específicos.

Esta constituye la primera versión del material y está sujeta a revisión. Si usted tiene sugerencias, nos sería de mucha utilidad conocerlas.

Deseamos agradecer la colaboración especial de Luis Roberto Sanint con respecto a la definición de temas, y a Marta Elena Carvajal en la transcripción.

- **MEDIDAS DE TENDENCIA CENTRAL**
  - Media
  - Mediana
  - Moda

- **MEDIDAS DE POSICION**
  - Mínimo y máximo
  - Percentiles
  - Deciles
  - Quintiles
  - Cuartiles

- **MEDIDAS DE DISPERSION**
  - Rango
  - Rango intercuartil, desviación cuartílica
  - Desviación media
  - Desviación estandar - Varianza

- **MEDIDAS DE DISPERSION RELATIVA**
  - Coeficiente de variación
  - Coeficiente de desviación media
  - Coeficiente de desviación cuartílica

- **MEDIDAS DE ASIMETRIA**
  - Coeficiente de Pearson
  - Coeficiente de Bowley
  - Sesgo

- **MEDIDAS DE CONCENTRACION**
  (Apuntamiento)

- **MOMENTOS DE UNA DISTRIBUCION**

- **EFECTOS DE TRANSFORMACIONES LINEALES, SOBRE ALGUNAS MEDIDAS DE TENDENCIA CENTRAL Y VARIABILIDAD**

- **SUMA/RESTA DE VARIABLES ALEATORIAS**

- **COVARIANZA**

- **CORRELACION**

- **DISTRIBUCION NORMAL**

- **CALIFICACIONES ESTANDAR**

- **DISTRIBUCION NORMAL ESTANDAR**

## ESTIMACION DE PARAMETROS

- ESTIMACION PUNTUAL - Condiciones de los estimadores
- ESTIMACION POR INTERVALO
- INTERVALOS DE CONFIANZA PARA $\mu$
- ESTIMACION POR INTERVALOS DE LA DIFERENCIA DE DOS MEDIAS
- ESTIMACION POR INTERVALO DEL PRODUCTO/RAZON DE VARIABLES

## NUMEROS INDICES

## ALGUNAS CONSIDERACIONES SOBRE LA PRESENTACION DE DATOS

## ANALISIS EXPLORATORIO DE DATOS

- Proporciona la oportunidad, para aquellos que no estén familiarizados con análisis estadísticos, de comprender los resultados, al conocer que están reflejando.

- Tanto en el caso de monitoreo como en de evaluaciones más profundas constituye el primer paso.

- Pretende revelar estructuras simples y patrones presentes en los datos.

- Permite detectar errores e inconsistencias antes de avanzar a fases más complejas del análisis.

- Incluyen análisis gráfico, ordenamiento, cálculo de medidas de posición, tendencia central y de dispersión (variabilidad), identificación de puntos extraños, posibles transformaciones para adecuado manejo de datos y detectar tendencias.

## ANALISIS GRAFICO

- Si se aprecian patrones especiales, se sugieren posibles estrategias de análisis y no solo el resumen escueto de resultados.

- Se pueden intuir relaciones especiales entre variables que ameriten una exploración profunda.

- Permite apreciar si los datos cumplen con los requerimientos teóricos de análisis y si no, las medidas remediales.

## MEDIDAS DE TENDENCIA CENTRAL

Indices de localización central, empleados en la descripción de las distribuciones de frecuencia. El centro de una distribución puede ser definida de diferentes maneras.

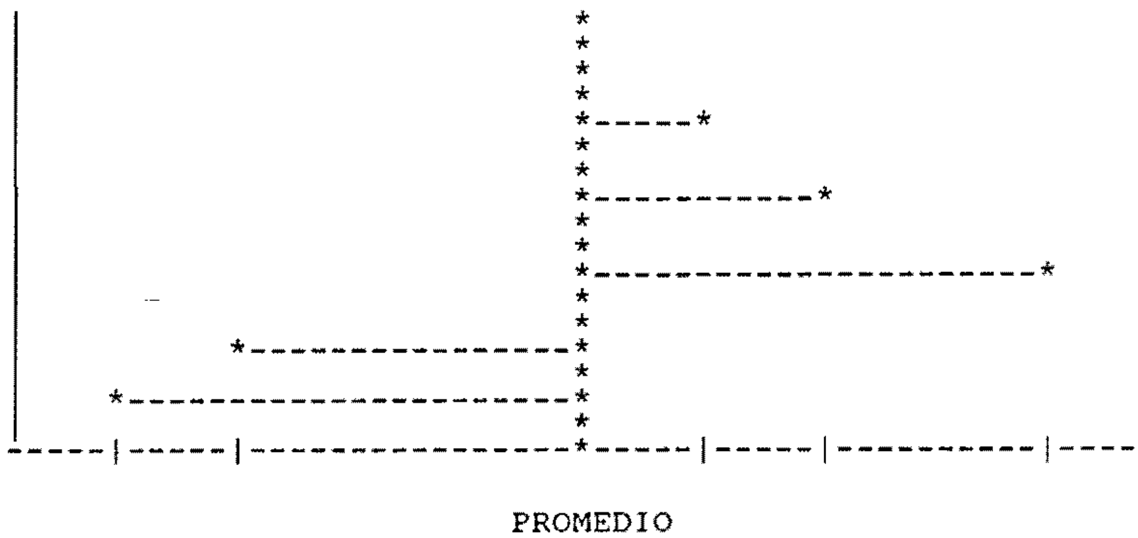
## MEDIA ARITMETICA

$$\overline{X} = ( X_1 + X_2 + \dots X_n ) / n$$

$$\overline{X} = ( \Sigma X_i ) / n$$

**CARACTERISTICAS :**

1.




desviación: $X_i - \overline{X}$

$\Sigma$ desviaciones $=$ $\Sigma \left( X_i - \overline{X} \right) = 0$

La media equilibra los puntajes superiores e inferiores a ella.

2. La media es muy sensible a las mediciones extremas cuando estas medidas no están equilibradas a ambos lados.

Ejemplos:

3, 5, 7, 9, 11 $\overline{X} = 7$

3, 5, 7, 9, 11, 25 $\overline{X} = 10$

3. La suma de los cuadrados de las desviaciones con respecto a la media es menor que si se toma otro valor como punto de referencia.

| | X | $(X-2)^2$ | $(X-20)^2$ | $(X-8)^2$ | $(X-9.2)^2$ |
|---|---|---|---|---|---|
| | 2 | 0 | 324 | 36 | 51.84 |
| | 4 | 4 | 256 | 16 | 27.04 |
| | 8 | 36 | 144 | 0 | 1.44 |
| | 12 | 100 | 64 | 16 | 7.84 |
| | 20 | 324 | 0 | 144 | 116.64 |
| $\Sigma$ | 46 | 464 | 788 | 212 | 204.8 |

$\overline{X}$ - 9.2

El método de localización de la media mediante la suma de los cuadrados de las desviaciones se llama " mínimos cuadrados ".

## LA MEDIANA

El valor real o potencial por debajo del cual se encuentra el 50% de los datos.

Si n, el número de datos es par : $p = n / 2$

$q = n / 2 + 1$

mediana : $( X_p + X_q ) / 2$

**Ejemplo 1**

4, 6, 8, 10

$n = 4$

$p = 2$

$q = 3$

Mediana : $( X_2 + X_3 ) / 2 = ( 6 + 8 ) / 2 = 7$

Si n, el número de datos es impar : $r = ( n + 1 ) / 2$

mediana : $X_r$

**Ejemplo 2**

4, 6, 8

$n = 3$

$r = ( 3 + 1 ) / 2 = 2$

mediana: $X_2 = 6$

**CARACTERISTICA**

Insensible a extremos.

No es obtenida con todos los datos.

No es única.

## LA MODA

Es el valor más frecuente. Se obtiene por inspección.

En algunos casos hay 2 valores más frecuentes, o varios, en cuyo caso se habla de distribución bimodal o multimodal respectivamente.

Ejemplo 1 : 2, 3, 3, 3, 4, 6 Moda = 3

Ejemplo 2 : 2, 3, 3, 4, 6, 6 Modas = 3,6

Ejemplo 3 : 2, 3, 4, 5, 6 Moda : no hay

## CARACTERISTICAS

No siempre puede determinarse.

Si se hace sobre datos agrupados puede cambiar según la definición de los límites de clase.

RESUMEN DE LAS PRINCIPALES CARACTERISTICAS DE LOS CINCO PROMEDIOS

| *Características* | *Media aritmética* ($\bar{X}$) | *Mediana* ($M_d$) | *Moda* ($M_0$) | *Media geométrica* ($G$) | *Media armónica* ($H$) |
|---|---|---|---|---|---|
| Cálculo: basado en | Cada valor | Valor central | Valor con la mayor frecuencia | Cada valor | Cada valor |
| Afectada por valores extremos | La más afectada | No (afectada solamente por elementos) | No | Menos que $\bar{X}$ | Menos que $G$ (más ponderación a los números más pequeños) |
| Manipulación algebraica | Sí: $\bar{X} = \frac{\sum X}{n}$ | No (promedio posicional, valor interpolado en muchos casos) | No (promedio concentrado, cuatro métodos para datos agrupados) | Sí: $G = \sqrt[n]{X_1 \cdot X_2 \cdots X_n}$ | Sí: $H = \frac{n}{\sum (1/X)}$ |
| Propiedades matemáticas | $\sum (X - \bar{X}) = 0$ $\sum (X - \bar{X})^2$ is es mínimo | $\sum (X - M_d)$ es mínimo (desdeñando signos) | — | $\sum (\log X - \log G) = 0$ | — |
| Aplicación con clases abiertas | Indeterminada | Determinada | Determinada | Indeterminada | Indeterminada |
| Comparación de respuestas para los mismos datos | Mayor que $G$ y $H$ | Entre $\bar{X}$ y $M_0$ | Puede ser mayor o menos que $M_d$ y $\bar{X}$ | Mayor que $H$, pero menor que $\bar{X}$ | Menor que $G$ y $\bar{X}$ |
| Tipo de datos preferidos | Mayoria de tipos | Valor central es típico, excluyendo extremos | Datos con distinta tendencia central | Razones, tasas y progresión geométrica | Ciertas tasas que pueden ser expresadas recíprocamente |

## MEDIDAS DE POSICION

Son medidas que indican la posición relativa de una calificación.

Para ello se parte de ordenar los datos ascendentemente y calcular la frecuencia acumulada, en términos absolutos y relativos.

El ordenamiento permite ver los extremos; **mínimo y máximo.**

Se define como **Rango Percentil,** el porcentaje de los casos que alcanzó valores menores que el citado.

A su vez, la calificación asociada con un percentil dado es el valor por debajo del cual esté acumulado el porcentaje señalado.

$$\text{Rango Percentil} = (\text{Frecuencia Acumulada} * 100) / N$$

$$\text{Frec. Acumulada} = (\text{Rango Percentil} * N) / 100$$

Algunos percentiles son muy utilizados, entre ellos están:

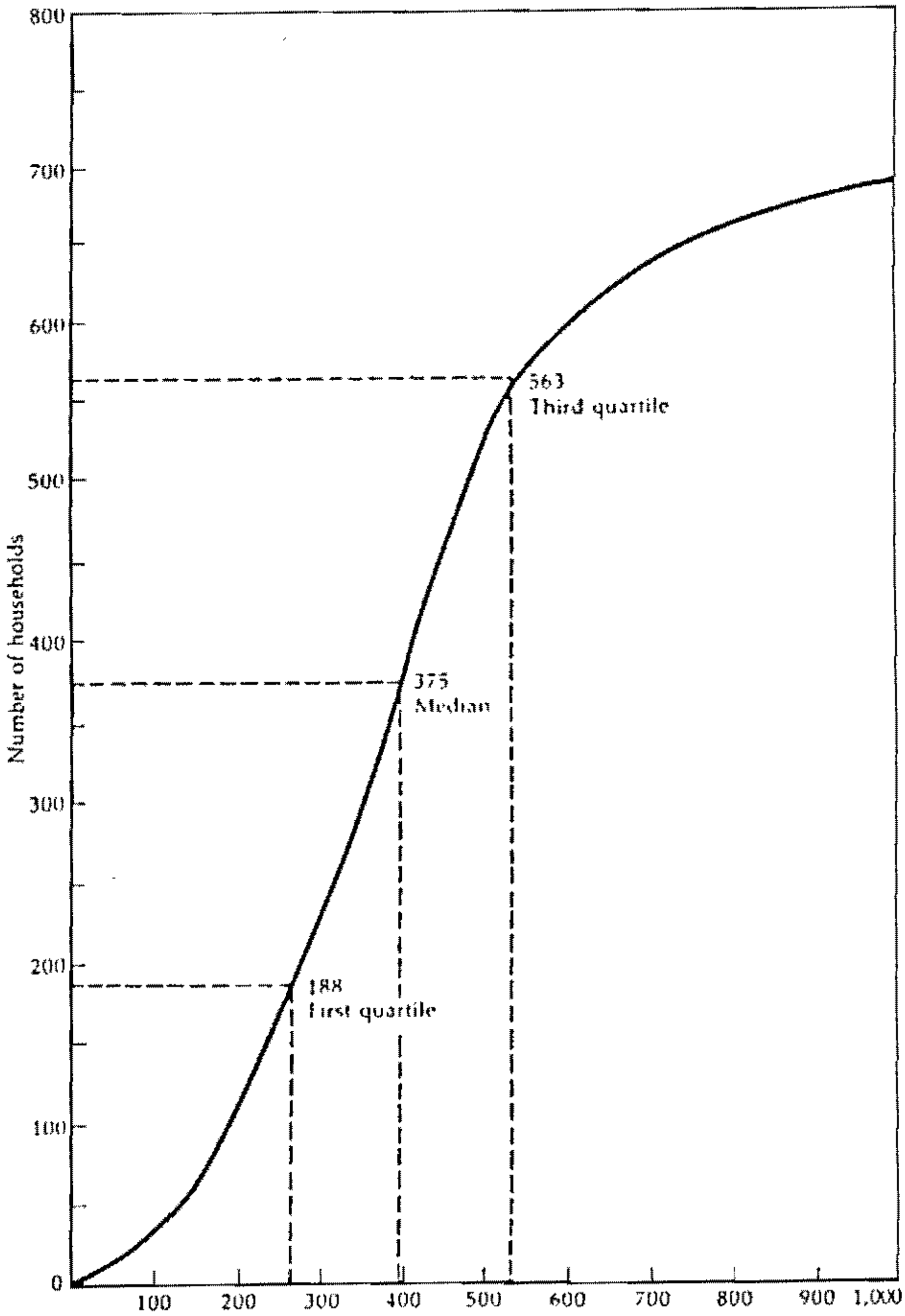
Percentil 10 = primer decil

Percentil 20 = segundo decil = primer quintil

Percentil 25 = primer cuartil

Percentil 50 = Mediana = segundo cuartil

Percentil 75 = tercer cuartil


800
700
600
500
400
300
200
100
0
Number of households
563
Third quartile
375
Median
188
First quartile
100
200
300
400
500
600
700
800
900
1,000
Income

**EJEMPLO**

Con el propósito de establecer el efecto de un suplemento alimenticio para ganado bobino, se registraron los datos de aumento de peso en un lote de terneros seleccionados lo más homogeneamente posible en términos de edad, peso, tamaño y raza.

El suplemento fue suministrado durante un periodo de 10 meses a un grupo de 100 terneros de raza Cebú, los cuales estaban bajo pastoreo.

Los valores obtenidos, en kilogramos, son los siguientes :

| OBS | RAN5 | DESVIACION | DESV CUADRATICA | DESV CUBICA | DESV A LA CUARTA |
|---|---|---|---|---|---|
| 1 | 118.651 | -27.2149 | 740.653 | -20156.81 | 548566.17 |
| 2 | 125.279 | -20.5872 | 423.832 | -8725.51 | 179633.55 |
| 3 | 125.477 | -20.3894 | 415.726 | -8476.39 | 172828.06 |
| 4 | 128.790 | -17.0763 | 291.600 | -4979.46 | 85030.85 |
| 5 | 128.798 | -17.0686 | 291.337 | -4972.71 | 84877.23 |
| 6 | 129.262 | -16.6045 | 275.711 | -4578.06 | 76016.57 |
| 7 | 130.128 | -15.7383 | 247.694 | -3898.29 | 61352.39 |
| 8 | 130.233 | -15.6332 | 244.398 | -3820.73 | 59730.44 |
| 9 | 130.417 | -15.4494 | 238.684 | -3687.53 | 56970.06 |
| 10 | 131.508 | -14.3578 | 206.147 | -2959.82 | 42496.63 |
| 11 | 131.951 | -13.9157 | 193.646 | -2694.71 | 37498.71 |
| 12 | 132.335 | -13.5316 | 183.104 | -2477.69 | 33527.04 |
| 13 | 132.493 | -13.3730 | 178.837 | -2391.59 | 31982.82 |
| 14 | 132.666 | -13.2001 | 174.242 | -2300.00 | 30360.19 |
| 15 | 132.858 | -13.0087 | 169.227 | -2201.43 | 28637.79 |
| 16 | 133.208 | -12.6587 | 160.242 | -2028.46 | 25677.63 |
| 17 | 133.626 | -12.2400 | 149.819 | -1833.79 | 22445.64 |
| 18 | 133.784 | -12.0823 | 145.982 | -1763.80 | 21310.73 |
| 19 | 134.408 | -11.4585 | 131.298 | -1504.47 | 17239.04 |
| 20 | 134.804 | -11.0625 | 122.378 | -1353.80 | 14976.38 |
| 21 | 135.057 | -10.8089 | 116.832 | -1262.83 | 13649.76 |
| 22 | 135.348 | -10.5186 | 110.641 | -1163.80 | 12241.53 |
| 23 | 137.198 | -8.6683 | 75.140 | -651.34 | 5646.04 |
| 24 | 137.711 | -8.1549 | 66.502 | -542.32 | 4422.55 |
| 25 | 137.912 | -7.9542 | 63.270 | -503.26 | 4003.07 |
| 26 | 138.054 | -7.8121 | 61.029 | -476.76 | 3724.50 |
| 27 | 138.251 | -7.6154 | 57.994 | -441.65 | 3363.33 |
| 28 | 138.359 | -7.5068 | 56.353 | -423.03 | 3175.63 |
| 29 | 138.511 | -7.3556 | 54.105 | -397.97 | 2927.31 |
| 30 | 138.601 | -7.2649 | 52.778 | -383.43 | 2785.54 |
| 31 | 138.735 | -7.1310 | 50.851 | -362.62 | 2585.84 |
| 32 | 139.675 | -6.1909 | 38.327 | -237.28 | 1468.95 |
| 33 | 140.160 | -5.7063 | 32.561 | -185.80 | 1060.25 |
| 34 | 140.469 | -5.3971 | 29.129 | -157.21 | 848.50 |
| 35 | 140.640 | -5.2267 | 27.318 | -142.79 | 746.30 |
| 36 | 141.003 | -4.8629 | 23.648 | -115.00 | 559.22 |
| 37 | 141.842 | -4.0245 | 16.197 | -65.18 | 262.33 |
| 38 | 141.942 | -3.9246 | 15.402 | -60.45 | 237.24 |
| 39 | 142.016 | -3.8503 | 14.824 | -57.08 | 219.77 |
| 40 | 142.088 | -3.7783 | 14.275 | -53.94 | 203.79 |

| OBS | RANS | DESVIACION | DESV CUADRATICA | DESV CUBICA | DESV A LA CUARTA |
|---|---|---|---|---|---|
| 41 | 142.144 | -3.72213 | 13.8543 | -51.567 | 191.94 |
| 42 | 142.443 | -3.42304 | 11.7172 | -40.109 | 137.29 |
| 43 | 142.480 | -3.38578 | 11.4635 | -38.813 | 131.41 |
| 44 | 143.472 | -2.39381 | 5.7303 | -13.717 | 32.84 |
| 45 | 143.496 | -2.37028 | 5.6182 | -13.317 | 31.56 |
| 46 | 143.519 | -2.34758 | 5.5111 | -12.938 | 30.37 |
| 47 | 144.441 | -1.42565 | 2.0325 | -2.898 | 4.13 |
| 48 | 144.552 | -1.31452 | 1.7280 | -2.271 | 2.99 |
| 49 | 144.695 | -1.17130 | 1.3720 | -1.607 | 1.88 |
| 50 | 144.780 | -1.08593 | 1.1792 | -1.281 | 1.39 |
| 51 | 144.862 | -1.00440 | 1.0088 | -1.013 | 1.02 |
| 52 | 144.942 | -0.92402 | 0.8538 | -0.789 | 0.73 |
| 53 | 145.101 | -0.76551 | 0.5860 | -0.449 | 0.34 |
| 54 | 145.378 | -0.48850 | 0.2386 | -0.117 | 0.06 |
| 55 | 145.477 | -0.38912 | 0.1514 | -0.059 | 0.02 |
| 56 | 145.679 | -0.18774 | 0.0352 | -0.007 | 0.00 |
| 57 | 146.069 | 0.20324 | 0.0413 | 0.008 | 0.00 |
| 58 | 146.189 | 0.32276 | 0.1042 | 0.034 | 0.01 |
| 59 | 146.289 | 0.42254 | 0.1785 | 0.075 | 0.03 |
| 60 | 147.054 | 1.18751 | 1.4102 | 1.675 | 1.99 |
| 61 | 147.397 | 1.53048 | 2.3424 | 3.585 | 5.49 |
| 62 | 147.487 | 1.62035 | 2.6255 | 4.254 | 6.89 |
| 63 | 147.595 | 1.72883 | 2.9888 | 5.167 | 8.93 |
| 64 | 148.179 | 2.31283 | 5.3492 | 12.372 | 28.61 |
| 65 | 149.522 | 3.65596 | 13.3660 | 48.866 | 178.65 |
| 66 | 150.123 | 4.25691 | 18.1213 | 77.141 | 328.38 |
| 67 | 150.639 | 4.77314 | 22.7828 | 108.746 | 519.06 |
| 68 | 150.828 | 4.96220 | 24.6234 | 122.186 | 606.31 |
| 69 | 151.185 | 5.31839 | 28.2853 | 150.433 | 800.06 |
| 70 | 151.215 | 5.34854 | 28.6069 | 153.005 | 818.35 |
| 71 | 151.545 | 5.67830 | 32.2431 | 183.086 | 1039.62 |
| 72 | 151.687 | 5.82040 | 33.8770 | 197.178 | 1147.65 |
| 73 | 151.741 | 5.87432 | 34.5076 | 202.709 | 1190.78 |
| 74 | 152.078 | 6.21178 | 38.5862 | 239.689 | 1488.89 |
| 75 | 152.289 | 6.42250 | 41.2485 | 264.919 | 1701.44 |
| 76 | 153.301 | 7.43468 | 55.2745 | 410.948 | 3055.27 |
| 77 | 153.563 | 7.69648 | 59.2359 | 455.908 | 3508.89 |
| 78 | 154.123 | 8.25684 | 68.1755 | 562.914 | 4647.90 |
| 79 | 154.247 | 8.38063 | 70.2350 | 588.614 | 4932.96 |

| OBS | RANS | DESVIACION | DESV CUADRATICA | DESV CUBICA | DESV A LA CUARTA |
|---|---|---|---|---|---|
| 80 | 155.231 | 9.3649 | 8,.70 | 821.30 | 7691.37 |
| 81 | 155.659 | 9.7931 | 95.90 | 939.20 | 9197.72 |
| 82 | 156.571 | 10.7049 | 114.59 | 1226.72 | 13131.92 |
| 83 | 156.981 | 11.1152 | 123.55 | 1373.24 | 15263.79 |
| 84 | 157.281 | 11.4143 | 130.29 | 1487.12 | 16974.47 |
| 85 | 157.347 | 11.4804 | 131.80 | 1513.11 | 17371.04 |
| 86 | 157.786 | 11.9200 | 142.09 | 1693.67 | 20188.55 |
| 87 | 157.871 | 12.0044 | 144.11 | 1729.90 | 20766.45 |
| 88 | 158.924 | 13.0576 | 170.50 | 2226.35 | 29070.84 |
| 89 | 159.207 | 13.3406 | 177.97 | 2374.25 | 31673.94 |
| 90 | 159.244 | 13.3782 | 178.97 | 2394.35 | 32032.02 |
| 91 | 159.826 | 13.9593 | 194.86 | 2720.14 | 37971.27 |
| 92 | 160.491 | 14.6250 | 213.89 | 3128.14 | 45749.05 |
| 93 | 161.666 | 15.8001 | 249.64 | 3944.38 | 62321.67 |
| 94 | 161.919 | 16.0525 | 257.68 | 4136.44 | 66400.03 |
| 95 | 166.686 | 20.8196 | 433.46 | 9024.38 | 187884.09 |
| 96 | 169.583 | 23.7167 | 562.48 | 13340.14 | 316383.44 |
| 97 | 173.064 | 27.1981 | 739.73 | 20119.36 | 547207.58 |
| 98 | 173.828 | 27.9616 | 781.85 | 21861.81 | 611291.03 |
| 99 | 175.652 | 29.7856 | 887.18 | 26425.33 | 787095.06 |
| 100 | 195.757 | 49.8908 | 2489.10 | 124183.12 | 6195600.80 |

UNIVARIATE PROCEDURE

VARIABLE=RAN5

FREQUENCY TABLE

| | | PERCENTS | | | | PERCENTS | |
|---|---|---|---|---|---|---|---|
| VALUE | COUNT | CELL | CUM | VALUE | COUNT | CELL | CUM |
| 118.6513 | 1 | 1.0 | 1.0 | 141.0033 | 1 | 1.0 | 36.0 |
| 125.2791 | 1 | 1.0 | 2.0 | 141.8418 | 1 | 1.0 | 37.0 |
| 125.4769 | 1 | 1.0 | 3.0 | 141.9416 | 1 | 1.0 | 38.0 |
| 128.7899 | 1 | 1.0 | 4.0 | 142.016 | 1 | 1.0 | 39.0 |
| 128.7977 | 1 | 1.0 | 5.0 | 142.088 | 1 | 1.0 | 40.0 |
| 129.2617 | 1 | 1.0 | 6.0 | 142.1441 | 1 | 1.0 | 41.0 |
| 130.1279 | 1 | 1.0 | 7.0 | 142.4432 | 1 | 1.0 | 42.0 |
| 130.233 | 1 | 1.0 | 8.0 | 142.4805 | 1 | 1.0 | 43.0 |
| 130.4168 | 1 | 1.0 | 9.0 | 143.4724 | 1 | 1.0 | 44.0 |
| 131.5084 | 1 | 1.0 | 10.0 | 143.496 | 1 | 1.0 | 45.0 |
| 131.9506 | 1 | 1.0 | 11.0 | 143.5187 | 1 | 1.0 | 46.0 |
| 132.3347 | 1 | 1.0 | 12.0 | 144.4406 | 1 | 1.0 | 47.0 |
| 132.4932 | 1 | 1.0 | 13.0 | 144.5517 | 1 | 1.0 | 48.0 |
| 132.6662 | 1 | 1.0 | 14.0 | 144.6949 | 1 | 1.0 | 49.0 |
| 132.8575 | 1 | 1.0 | 15.0 | 144.7803 | 1 | 1.0 | 50.0 |
| 133.2076 | 1 | 1.0 | 16.0 | 144.8619 | 1 | 1.0 | 51.0 |
| 133.6262 | 1 | 1.0 | 17.0 | 144.9422 | 1 | 1.0 | 52.0 |
| 133.7839 | 1 | 1.0 | 18.0 | 145.1007 | 1 | 1.0 | 53.0 |
| 134.4077 | 1 | 1.0 | 19.0 | 145.3777 | 1 | 1.0 | 54.0 |
| 134.8038 | 1 | 1.0 | 20.0 | 145.4771 | 1 | 1.0 | 55.0 |
| 135.0574 | 1 | 1.0 | 21.0 | 145.6785 | 1 | 1.0 | 56.0 |
| 135.3476 | 1 | 1.0 | 22.0 | 146.0695 | 1 | 1.0 | 57.0 |
| 137.1979 | 1 | 1.0 | 23.0 | 146.189 | 1 | 1.0 | 58.0 |
| 137.7114 | 1 | 1.0 | 24.0 | 146.2888 | 1 | 1.0 | 59.0 |
| 137.912 | 1 | 1.0 | 25.0 | 147.0538 | 1 | 1.0 | 60.0 |
| 138.0542 | 1 | 1.0 | 26.0 | 147.3967 | 1 | 1.0 | 61.0 |
| 138.2509 | 1 | 1.0 | 27.0 | 147.4866 | 1 | 1.0 | 62.0 |
| 138.3594 | 1 | 1.0 | 28.0 | 147.5951 | 1 | 1.0 | 63.0 |
| 138.5107 | 1 | 1.0 | 29.0 | 148.1791 | 1 | 1.0 | 64.0 |
| 138.6014 | 1 | 1.0 | 30.0 | 149.5222 | 1 | 1.0 | 65.0 |
| 138.7352 | 1 | 1.0 | 31.0 | 150.1232 | 1 | 1.0 | 66.0 |
| 139.6754 | 1 | 1.0 | 32.0 | 150.6394 | 1 | 1.0 | 67.0 |
| 140.16 | 1 | 1.0 | 33.0 | 150.8284 | 1 | 1.0 | 68.0 |
| 140.4691 | 1 | 1.0 | 34.0 | 151.1846 | 1 | 1.0 | 69.0 |
| 140.6395 | 1 | 1.0 | 35.0 | 151.2148 | 1 | 1.0 | 70.0 |

UNIVARIATE PROCEDURE

VARIABLE=RAN5

FREQUENCY TABLE (CONT.)

| | | PERCENTS | | | | PERCENTS | |
|---|---|---|---|---|---|---|---|
| VALUE | COUNT | CELL | CUM | VALUE | COUNT | CELL | CUM |
| 151.5446 | 1 | 1.0 | 71.0 | 157.7862 | 1 | 1.0 | 86.0 |
| 151.6866 | 1 | 1.0 | 72.0 | 157.8707 | 1 | 1.0 | 87.0 |
| 151.7406 | 1 | 1.0 | 73.0 | 158.9239 | 1 | 1.0 | 88.0 |
| 152.078 | 1 | 1.0 | 74.0 | 159.2069 | 1 | 1.0 | 89.0 |
| 152.2888 | 1 | 1.0 | 75.0 | 159.2444 | 1 | 1.0 | 90.0 |
| 153.3009 | 1 | 1.0 | 76.0 | 159.8256 | 1 | 1.0 | 91.0 |
| 153.5627 | 1 | 1.0 | 77.0 | 160.4912 | 1 | 1.0 | 92.0 |
| 154.1231 | 1 | 1.0 | 78.0 | 161.6663 | 1 | 1.0 | 93.0 |
| 154.2469 | 1 | 1.0 | 79.0 | 161.9187 | 1 | 1.0 | 94.0 |
| 155.2311 | 1 | 1.0 | 80.0 | 166.6859 | 1 | 1.0 | 95.0 |
| 155.6593 | 1 | 1.0 | 81.0 | 169.5829 | 1 | 1.0 | 96.0 |
| 156.5711 | 1 | 1.0 | 82.0 | 173.0643 | 1 | 1.0 | 97.0 |
| 156.9814 | 1 | 1.0 | 83.0 | 173.8278 | 1 | 1.0 | 98.0 |
| 157.2805 | 1 | 1.0 | 84.0 | 175.6519 | 1 | 1.0 | 99.0 |
| 157.3466 | 1 | 1.0 | 85.0 | 195.7571 | 1 | 1.0 | 100.0 |

UNIVARIATE PROCEDURE

VARIABLE=RAN5

MOMENTS

| | | | |
|---|---|---|---|
| N | 100 | SUM WGTS | 100 |
| MEAN | 145.8662 | SUM | 14586.62 |
| STD DEV | 12.27479 | VARIANCE | 150.6705 |
| SKEWNESS | 0.868212 | KURTOSIS | 2.009364 |
| USS | 2142613 | CSS | 14916.38 |
| CV | 8.4151 | STD MEAN | 1.227479 |
| T:MEAN=0 | 118.834 | PROB>\|T\| | 0.0001 |
| NUM ¬= 0 | 100 | NUM > 0 | 100 |
| M(SIGN) | 50 | PROB>\|M\| | 0.0001 |
| SGN RANK | 2525 | PROB>\|S\| | 0.0001 |

QUANTILES(DEF=5)

| | | | |
|---|---|---|---|
| 100% MAX | 195.7571 | 99% | 185.7045 |
| 75% Q3 | 152.7948 | 95% | 168.1344 |
| 50% MED | 144.8211 | 90% | 159.535 |
| 25% Q1 | 137.9831 | 10% | 131.7295 |
| 0% MIN | 118.6513 | 5% | 129.0297 |
| | | 1% | 121.9652 |
| RANGE | 77.10578 | | |
| Q3-Q1 | 14.81175 | | |
| MODE | 118.6513 | | |

EXTREMES

| LOWEST | OBS | HIGHEST | OBS |
|---|---|---|---|
| 118.6513( | 81) | 169.5829( | 16) |
| 125.2791( | 62) | 173.0643( | 72) |
| 125.4769( | 69) | 173.8278( | 45) |
| 128.7899( | 37) | 175.6519( | 73) |
| 128.7977( | 66) | 195.7571( | 93) |

## MEDIDAS DE DISPERSION ABSOLUTA

Para describir más completamente una distribución se necesita información relativa a la dispersión con respecto a la medida de tendencia central.

Este índice indica la distancia en la escala de calificación.

## RANGO

Distancia escalar entre la mayor y la menor de las calificaciones.

## CARACTERISTICAS

Inestable por su sensibilidad a extremos.

**RANGO INTERCUARTIL**

Calificación correspondiente al percentil 75 = $Q_3$

Calificación correspondiente al percentil 25 = $Q_1$

Rango intercuartil = $Q_3 - Q_1$

**DESVIACION CUARTILICA**

Se calcula a partir del rango intercuartil

$DC = ( Q_3 - Q_1 ) / 2$

## CARACTERISTICAS

1. No permite hacer una interpretación precisa de una calificación en la distribución.

2. Con respecto a esta medida no se han hecho desarrollos analíticos superiores.

3. Insensible a valores extremos.

## DESVIACION MEDIA

Desviación con respecto a la media = $X_i$ - Media

DM = ( $\Sigma$ |desviaciones con respecto a la media| ) / N

## CARACTERISTICAS

Representa una medida muy práctica, pero debido al manejo incómodo del " valor absoluto " se rechaza.

Es mínima si se considera con respecto a la mediana.

En el ejemplo que estamos considerando

Rango = 77.11

$Q_3$ = 152.79

$Q_1$ = 137.98

Rango intercuartil = 14.81

Desviación cuartílica = 7.40

Desviación media = 9.33

## DESVIACION ESTANDAR Y VARIANZA

Se toma la desviación con respecto a la media, se eleva al cuadrado (con ello se elimina el problema de los signos) y se calcula el promedio, este valor se llama varianza.

```
|                           *
|                           *
|                           *
|                           *
|                           *-----*
|                           *
|                           *
|                           *-----------*
|                           *
|                           *
|                           *----------------------*
|                           *
|                           *
|          *-- DESVIACION --*
|                           *
|    *----------------------*
|                           *
-----|-----|----------------*-----|-----|----------|----

                        PROMEDIO
```

$$S^2 = \Sigma ( X_i - \bar{X} )^2 / N$$

$$S^2 = ( \Sigma X_i^2 ) / N - \bar{X}^2$$

Note que:

$$\Sigma X_i^2 \neq ( \Sigma X_i )^2$$

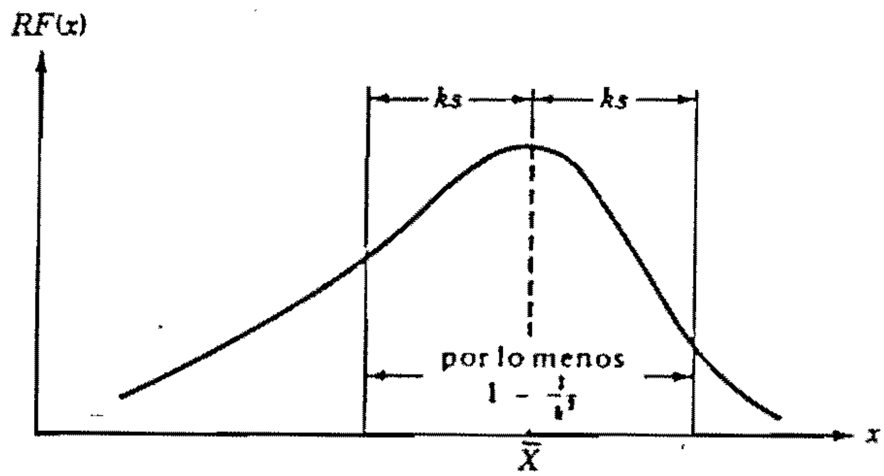
La raíz cuadrada de la varianza es la desviación estándar la cual está expresada en unidades iguales a las de la variable.

El hecho de que la varianza sea mínima cuando se calcula con respecto a la media, nos permite suponer que si sobre cierta población hacemos una predicción igual a su promedio, estaremos cometiendo el mínimo error. Esto hace que las medidas sean ampliamente usadas en conjunto.

## SIGNIFICADO PRACTICO DE LA DESVIACION STANDAR

## DESIGUALDAD DE TCHEBYSHEFF :

Dado un número $K \geq 1$ y un conjunto de n observaciones $Y_1$ , $Y_2$ , $Y_n$ , por lo menos $1 - 1/K^2$ de ellas se encuentran dentro de K desviaciones estándar de la media.




al menos $1 - 1/K^2$

$$P ( \mu - K\sigma \leq X \leq \mu + K\sigma ) \geq 1 - K^{-2}$$

**Ejemplo**

$\bar{Y} = 75$ $\quad S_y^{\,2} = 100$ $\quad n = 25$

Describa la distribución

$(\bar{Y} \pm 2S) = 75 \pm 2 * 10 = (55 - 95)$ : 3/4 partes de los datos.

$(\bar{Y} \pm 3S) = 75 \pm 3 * 10 = (45 - 105)$ : 8/9 partes de los datos

Lo anterior permite acercarse a la $S^2$ si conocemos el valor del rango.

RESUMEN DE LAS PRINCIPALES CARACTERISTICAS DE LA DISPERSION ABSOLUTA

| *Características* | *Recorrido* ($R$) | *Desviación cuartílica* (Q.D.) | *Desviación media* (A.D.) | *Desviación estándar* ($s$) |
|---|---|---|---|---|
| Cálculo: basado en | Valores mínimo y máximo | $Q_1$ y $Q_3$ | Cada valor | Cada valor |
| Afectado por valores extremos | El más afectado | No por los valores menores que $Q_1$ ni mayores que $Q_3$ | Afectado por cada valor | Afectado por cada valor |
| Grado de precisión como una medida de dispersión | Tosca estimación | Sólo mejor que el recorrido | Buena, pero solamente mide desviaciones absolutas con respecto a la media o a la mediana. | Excelente, mide las desviaciones al cuadrado con respecto a la media |
| Ventajas matemáticas | Fácil de calcular | Puede ser usada para medir distribuciones asimétricas | Más fácil de calcular que la desviación estándar | Difícil de calcular, pero apropiada para cálculos matemáticos adicionales |

## DISPERSION RELATIVA

Si dos conjuntos de valores están siendo comparados, los valores de su dispersión son claros, solo cuando sus promedios son similares y las unidades de medida son iguales.

Aún, si las unidades de medida fueran las mismas, la diferencia en los promedios dificulta la comparación por lo tanto, una medida de dispersión debe estandarizarse. Este hecho crea el coeficiente de variación.

$$CV = S/\bar{X}$$

el cual es adimensional y cumple con los objetivos.

Hay otras medidas similares

Coeficiente Recorrido = Rango/ [Max + Min] / 2

Coeficiente Desviación Media = Desv. media / $\overline{X}$, Md

Coeficiente Desviación Cuartílica = $( Q_3 - Q_1 ) / ( Q_3 + Q_1 )$

Retomando el ejemplo tenemos:

Coeficiente de variación = 0.084

Coeficiente de recorrido = 0.49

Coeficiente de desviación media ( respecto a $\overline{X}$ ) = 0.064

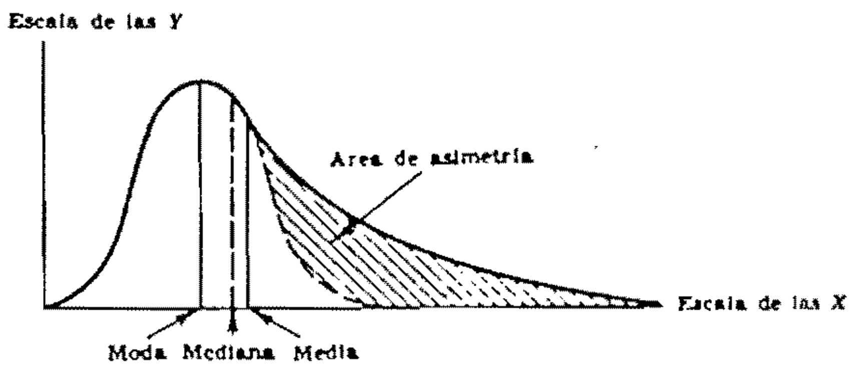
Coeficiente de desviación media ( respecto a Md ) = 0.064

Coeficiente de desviación cuartílica = 0.051

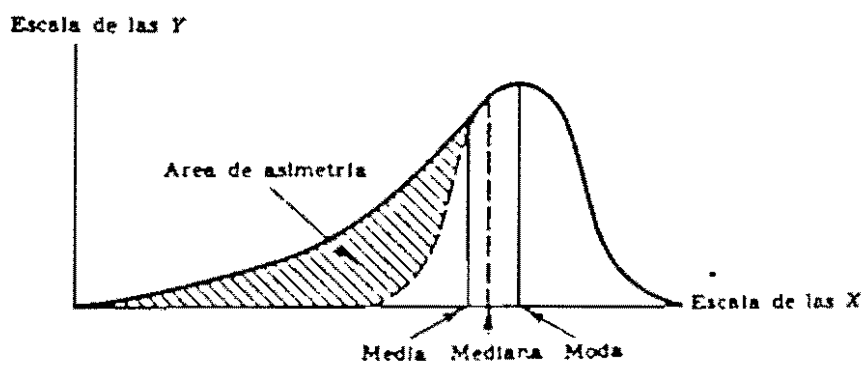
**MEDIDAS DE SIMETRIA**

Si al tomar la medida de tendencia central como un eje de referencia, la distribución de los datos a izquierda y a derecha es igual, hablamos de distribución simétrica. Si no es así, decimos que tiene sesgo.

Sesgo positivo indica que la mayoría de los valores están al lado izquierdo, pero hay algunos valores al lado derecho que pueden tener magnitud extrema.




Sesgo negativo : la mayoría de los valores están al lado derecho, pero hay algunos valores al lado izquierdo que pueden tener magnitud extrema.



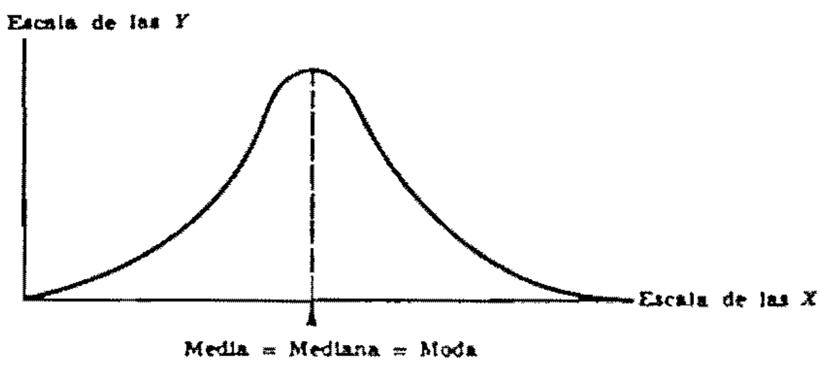
## ASIMETRIA

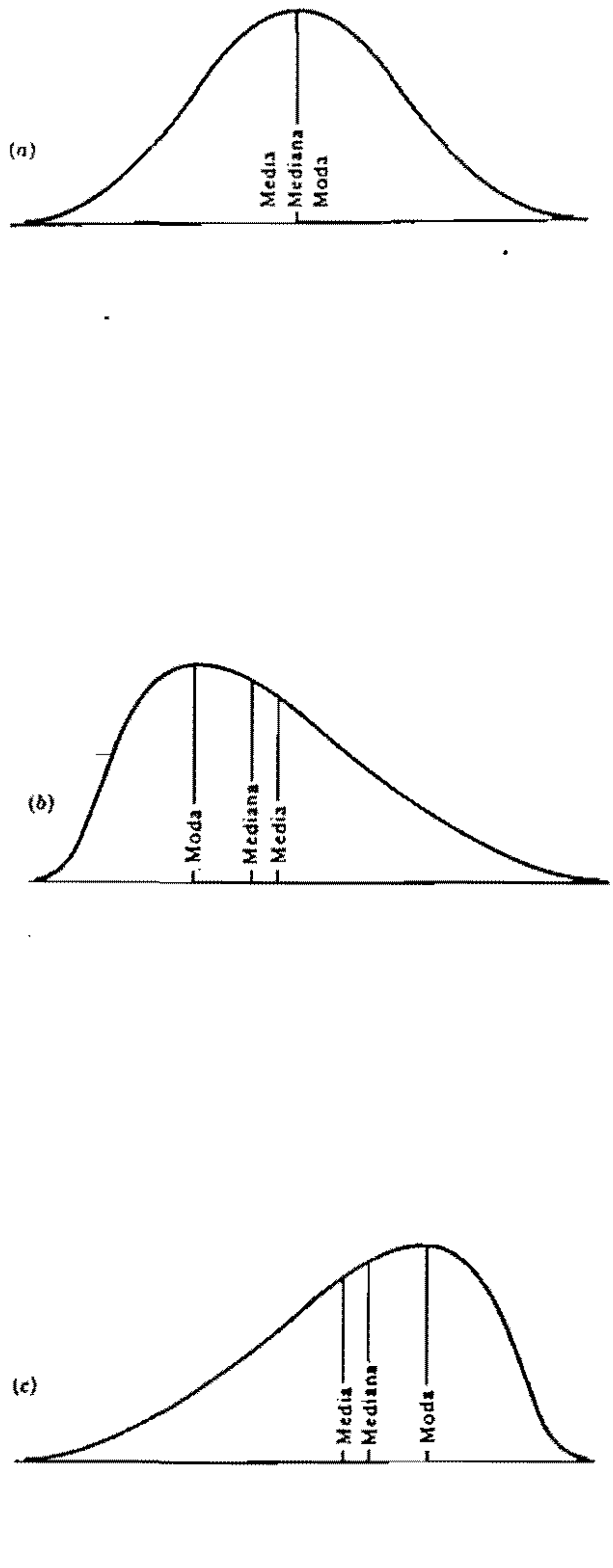
Si la distribución es simétrica $\overline{X}$ = Md = Mo si no es así empieza a presentarse asimetría y sería mayor en la medida en que ellas difieran.

$\overline{X} \leq Md \leq Mo$ si la asimetría es negativa

$Mo \leq Md \leq \overline{X}$ si la asimetría es positiva

I. Distribución simétrica






La posición de la media, la mediana y la moda en (*a*) una distribución simétrica, (*b*) una distribución de asimetría positiva, y (*c*) una distribución de asimetría negativa.

**MEDIDAS**

1. Pearson $= ( \overline{X} - Mo ) / S$

2. A. Bowley $= [ ( Q_3 - Q_2 ) - ( Q_2 - Q_1 ) ] / ( Q_3 - Q_1 )$

   $-1 \leq S \leq 1$ $|0.1|$: moderado $|0.3|$: alto

3. Sesgo $= \Sigma ( X1 - \overline{X} )^3 / n$

**Ejemplo:**

Para los datos de :
Coeficiente de Pearson = No se puede determinar
Coeficiente de Bowley = 0.0766
Coeficiente de Sesgo = 0.868212

**APUNTAMIENTO O KURTOSIS**

Es el cuarto recurso para identificar una distribución y muestra el grado de concentración de los datos.

Se mide por el coeficiente de Kurtosis. ( Ver Tabla de Momentos )

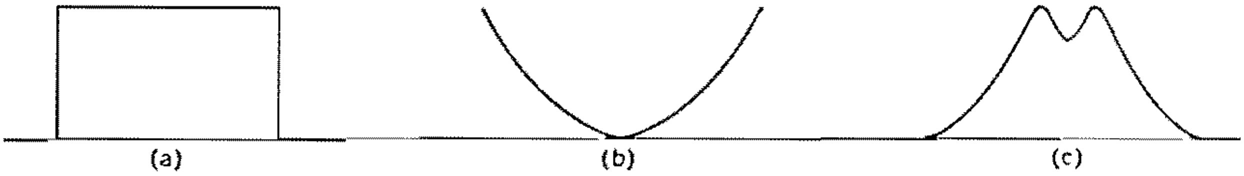
## MOMENTOS

Sea $d_i$ = desviación con respecto a la media.

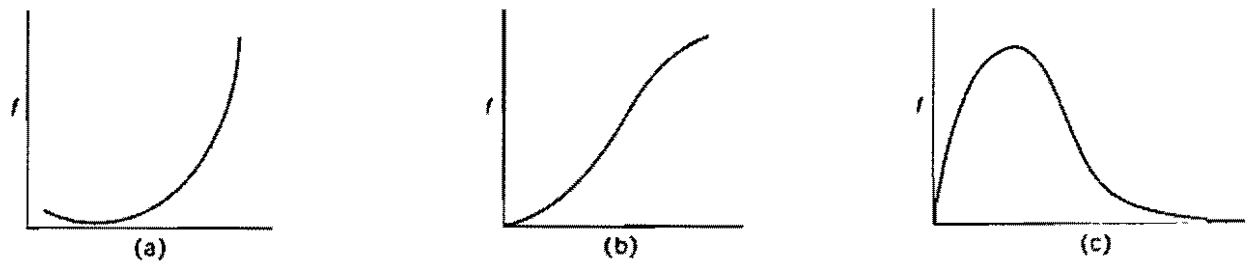
$$d_i = X_i - \bar{X}$$

el momento k = $\Sigma\, d_i^k / n$

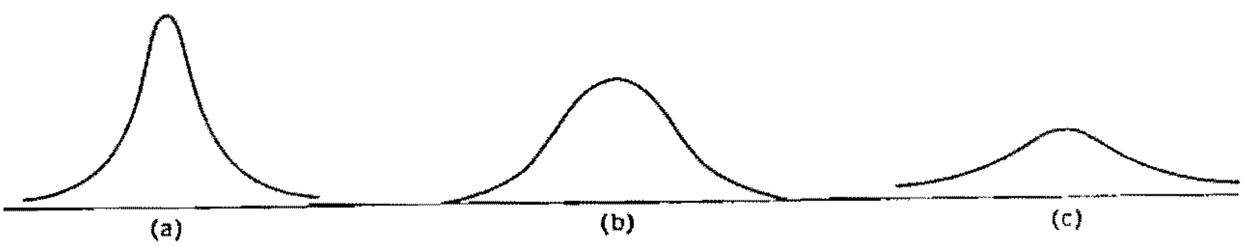
| momento | fórmula | uso |
|---|---|---|
| $m_1$ | $\Sigma\, d_i /n = 0$ | |
| $m_2$ | $\Sigma\, d_i^2/n$ | Varianza |
| $m_3$ | $\Sigma\, d_i^3/n$ | Sesgo = $m_3 / S^3$ |
| $m_4$ | $\Sigma\, d_i^4/n$ | Kurtosis = $m_4 / S^4$ |




Representación de varios polígonos de frecuencias simetricos anormales.




Representación de polígonos de frecuencia asiméticos.




Tres formas de distribuciones normales: (a) leptocúrtica, (b) mesocúrtica y (c) platicúrtica.

UNIVARIATE PROCEDURE

VARIABLE=RAN5

```
STEM LEAF                      #            BOXPLOT
  19 6                         1               0
  19
  18
  18
  17 6                         1               0
  17 034                       3               |
  16 7                         1               |
  16 0022                      4               |
  15 56777788999              11               |
  15 001111222223444          15            +-----+
  14 55555555666677788        17            *--+--*
  14 0001122222223344         16            |     |
  13 555788888999             12            +-----+
  13 0002222333444            13               |
  12 55999                     5               |
  12                                           |
  11 9                         1               |
     ----+----+----+----+
 MULTIPLY STEM.LEAF BY 10**+1
```

**EFECTO DE TRANSFORMACIONES LINEALES DE VARIABLES SOBRE ALGUNAS MEDIDAS DE TENDENCIA CENTRAL Y VARIABILIDAD**

| $X$ | $Y_1 = X + 3$ | $Y_2 = 2X$ | $Y_3 = 2X + 3$ |
|---|---|---|---|
| 2 | 5 | 4 | 7 |
| 3 | 6 | 6 | 9 |
| 4 | 7 | 8 | 11 |
| 5 | 8 | 10 | 13 |
| 6 | 9 | 12 | 15 |
| 7 | 10 | 14 | 17 |
| 8 | 11 | 16 | 19 |
| $\overline{X} = 5$ | $\overline{Y}_1 = 8 = \overline{X} + 3$ | $\overline{Y}_2 = 10 = 2\overline{X}$ | $\overline{Y}_3 = 13 = 2\overline{X} + 3$ |
| $Md = 5$ | $8 = Md + 3$ | $10 = 2\ Md$ | $13 = 2\ Md + 3$ |
| $Rango = 6$ | 6 | $12 = 2R$ | $12 = 2R$ |
| $S^2 = 4$ | 4 | $16 = 2^2 * S^2$ | $16 = 2^2 * S^2$ |

## SUMA / RESTA DE VARIABLES ALEATORIAS

Si X, Y son dos variables aleatorias y

$Z = X \pm Y$

$\overline{Z} = \overline{X} \pm \overline{Y}$

$S^2_z = S^2_x + S^2_y \pm 2 \operatorname{cov}(X, Y)$

$\operatorname{Cov}(X, Y) = \Sigma (X_i - \overline{X})(Y_i - \overline{Y}) / n$

es la covarianza entre X, Y

Si la variable X es independiente de Y entonces cov (X,Y) = 0

pero Cov ( X, Y ) = 0 NO IMPLICA INDEPENDENCIA

La covarianza entre X, Y mide la influencia en la variación absoluta de una variable sobre la otra.

Usualmente se usa estandarizada para restringir sus valores al intervalo -1, 1.

La covarianza estandarizada recibe el nombre de correlación (r) y se calcula así:

$$r = \frac{\text{cov}(X, Y)}{S_x \, S_y}$$

$S_x$ = desviación estandar X

$S_y$ = desviación estandar Y

Posteriormente volveremos sobre este tema.

**EJEMPLO**

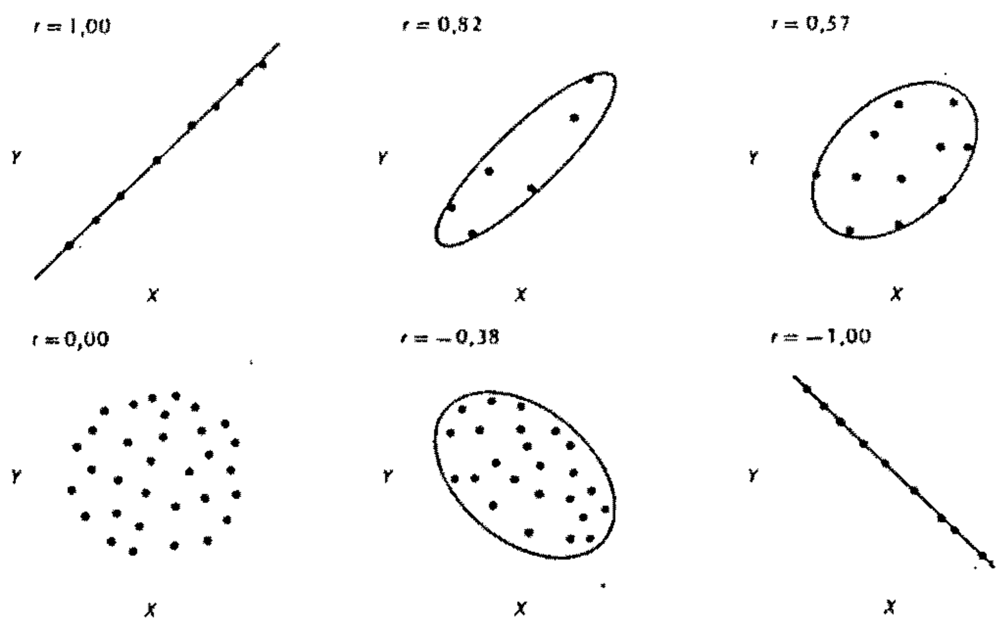
En lotes comerciales de habichuela el número promedio de larvas de mosca blanca por trifolio tiene las siguientes características:

| | | |
|---|---|---|
| Larvas pequeñas | $\overline{X} = 10.69$ | $S^2 = 72.7899$ |
| Larvas grandes | $\overline{X} = 11.55$ | $S^2 = 642.414$ |
| Total larvas | $\overline{X} = 22.24$ | $S^2 = 935.944$ |

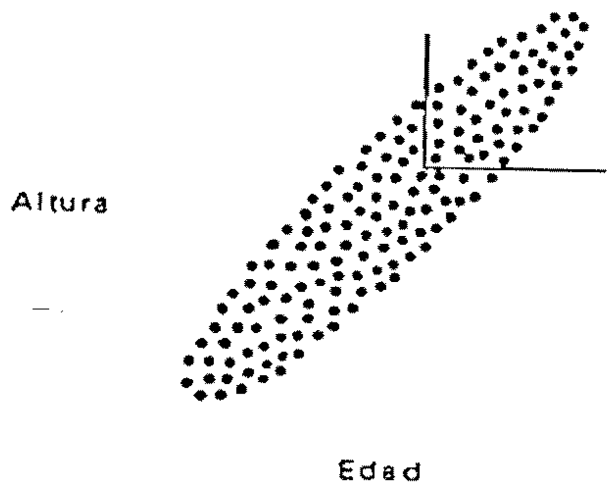
Con los datos anteriores puede verificarse que el promedio de la suma es la suma de los promedios. La varianza de la suma no es la suma de las varianzas, lo cual se presenta debido a la covarianza.

$935.944 = 72.7899 + 642.414 + 2.\ \mathrm{cov}\ (G, P)$

entonces $\mathrm{Cov}\ (G, P) = 110.37005$




Diagramas de dispersión que muestran varios grados de relación entre dos variables.



## FUNCION DE DISTRIBUCION DE PROBABILIDADES

La distribución de probabilidades ( Y ) de una variable aleatoria X, es una función de densidad.

Si para todos los valores **a** y **b**, la probabilidad de que X esté entre a y b está representada por el área bajo la curva de Y, entre a y b.

## DISTRIBUCION NORMAL

Sea $Z = (X - \mu) / \sigma$

$$Y = f(X) = \frac{1}{\sigma \sqrt{2\pi}} * \exp(-Z^2 / 2)$$

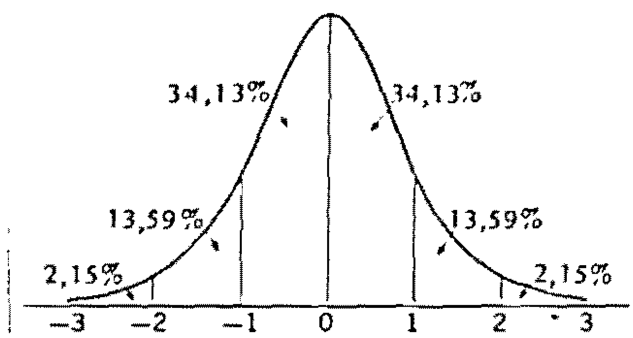
Y es una curva lisa y de forma de campana con propiedades muy especiales:

1. Dominio: R

2. $P(|X - \mu| < \sigma) = 68.27\ \%$

3. $P(|X - \mu| < 2\sigma) = 95.45\ \%$

4. $P(|X - \mu| < 3\sigma) = 99.73\ \%$



## CALIFICACIONES ESTANDAR

Para interpretar el valor de un dato, queremos en general tener algún tipo de referencia, por ejemplo.

Un percentil (% de valores menores que él). Otra alternativa es mirar su desviación con respecto a la media.

Pero, que tanto representa esta desviación en el conjunto de datos?

La solución sería comparar la desviación de este punto con la desviación estandar obteniendo un nuevo valor, Z que tiene la interesante característica de ser adimensional.

$$Z = X - \overline{X} / S$$

## QUE ES UNA DISTRIBUCION NORMAL ESTANDAR ?

La distribución normal estandar

1. Media $\mu = 0$

2. Varianza $\sigma^2 = 1$

3. Sesgo = 0 (simetría)

4. Kurtosis = 3

5. Por ser función de probabilidad área bajo la curva = 1
   y $P ( X \leq K) =$ Area bajo la curva hasta el punto K.

6. Cambios de concavidad en $Z = \pm 1$

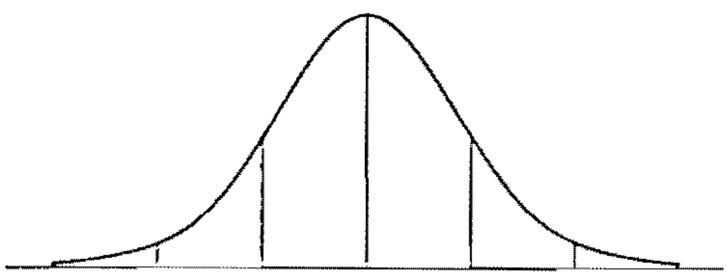
7. El eje X es asíntota a partir de + 3 y de - 3

Como Z es una transformación lineal de X ( se le resta una constante y el resultado se divide por otra) puede deducirse facilmente que

1. Sumatoria de Z = 0
2. Media de Z = 0
3. Sumatoria de $Z^2$ = N
4. Varianza de Z = 1

**CONCLUSION**

Si una variable tiene distribución normal, cualquiera de sus valores puede expresarse como un percentil encontrando su valor Z correspondiente y aplicando las propiedades de la distribución normal estándar.



| Calificaciones | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|---|---|---|---|---|---|---|---|
| Calificación $z$ | −3 | −2 | −1 | 0 | +1 | +2 | +3 |
| Rango percentil | ,13 | 2,28 | 15,87 | 50,00 | 84,13 | 97,72 | 99,87 |

Algunos valores importantes de $Z_c$

| $\alpha$ | CONFIABILIDAD % | $Z_c$ |
|---|---|---|
| 1 | 99 | 2.58 |
| 2 | 98 | 2.33 |
| 5 | 95 | 1.96 |
| 10 | 90 | 1.65 |

**NOTA**

Todo lo visto hasta ahora hace perfecta referencia a la descripción de una POBLACION.

A PARTIR DE AHORA, pensaremos en términos de subconjuntos de esa población llamados muestras.

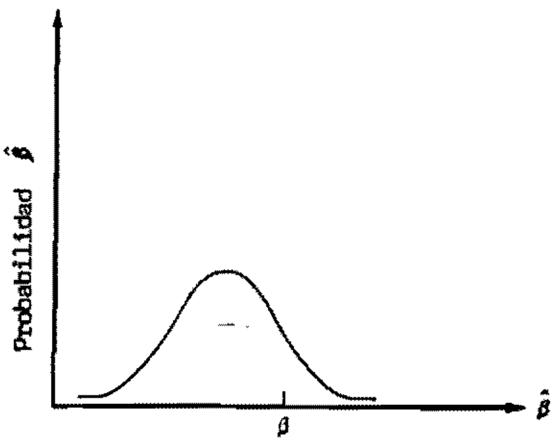
## ESTIMACION DE PARAMETROS
## ESTIMACION PUNTUAL

Si se está en posesión de una muestra y se desea conocer mediante ella lo que ocurre en la población necesitamos preguntarnos

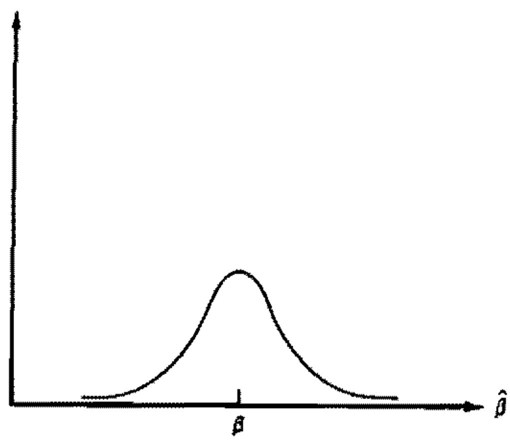
Hasta dónde los estadísticos que hemos presentado son buenos estimadores de los parámetros respectivos?

El mejor estadístico debe poseer ciertas propiedades como :

**INSESGAMIENTO :** La media de todas las posibles estimaciones sea el parámetro que se estima.

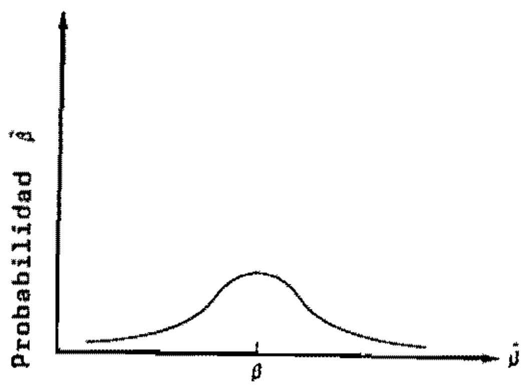



Estimador Sesgado

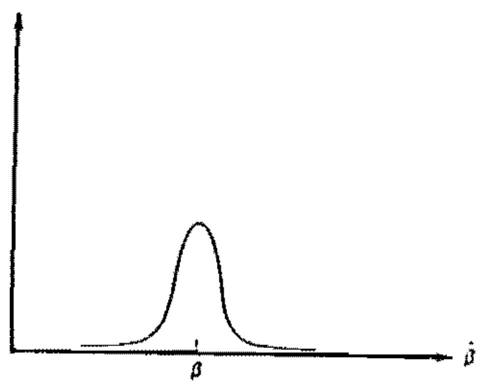



Estimador Insesgado

**VARIANZA PEQUEÑA (EFICIENCIA)** : De todas las alternativas aquella que sea más estable.

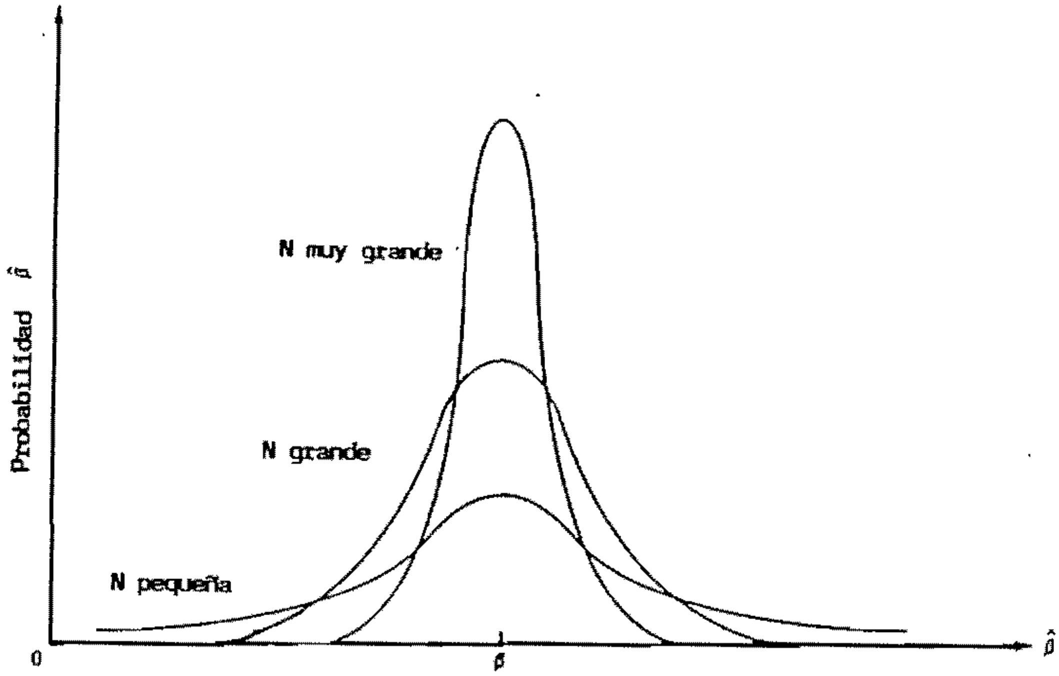



Estimador Ineficiente




Estimador Eficiente

**CONSISTENCIA** : El buen estadístico debe acercarse asintóticamente al parámetro en la medida en que aumenta el tamaño de la muestra.



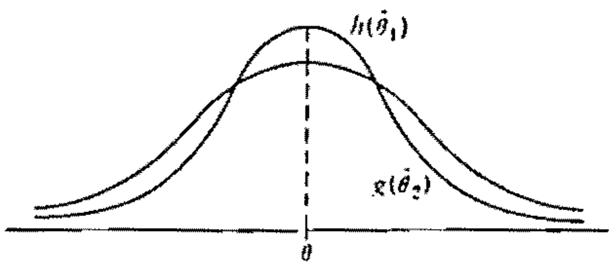

**FUNCION LINEAL DE DATOS : Sencillez de cálculo y uso total de los datos**

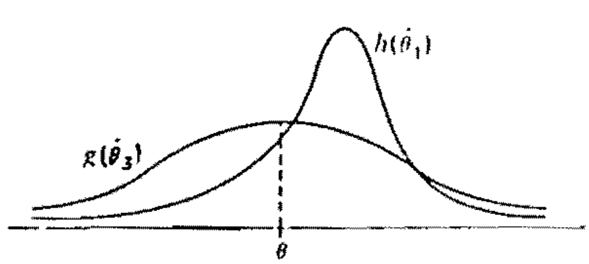
M.E.L.I. : Mejor estimador lineal insesgado

$\overline{X}$ : M.E.L.I.

$S^2$ : Sesgado, no es lineal

$S^2 : \Sigma ( X_i - \overline{X} )^2 / ( n - 1 )$ es un estimador insesgado de $\sigma^2$



**ESTIMACION POR INTERVALOS**

**INFERENCIA ESTADISTICA :**

Si a partir de una población con media $\mu$ y varianza $\sigma^2$ se toman varias muestras y para cada muestra se calcula el promedio X se tiene que :

1. Puede calcularse el promedio de los promedios

2. Y la varianza de los promedios

El teorema del LIMITE CENTRAL dice : si se extraen repetidamente muestras al azar de tamaño N, de una población normal, la distribución resultante de las medias será normal, con una media tendiendo a ser igual a la media de la población.

La varianza de las medias será $\sigma^2 / N$ y el valor $\sigma / \sqrt{N}$ (desviación estándar de la media) recibe el nombre de error estándar.

La ley de los grandes números establece que si N es grande ( $N \geq 30$ ) el teorema del LIMITE CENTRAL se cumple, aún cuando la población no sea normal.

**RESUMEN**

$\bar{X} \sim N ( \mu, \sigma^2 / N )$

Por lo tanto, para $\bar{X}$ puede construirse también un valor estándar:

$Z = ( \bar{X} - \mu ) / S_{\bar{x}}$

y calcularse probabilidades

## ACLARACION

Estimación no sesgada de la varianza

$$S^2 = \Sigma ( X_i - \bar{X} )^2 / ( N - 1 )$$

Error estándar = $S_{\bar{x}}$

$$S_{\bar{x}} = S / \sqrt{N}$$

$$S_{\bar{x}} = \sqrt{\Sigma ( X_i - \bar{X} )^2 / ( N - 1 ) \ N}$$

ILUSTRACION DE LA DISTRIBUCION EN EL MUESTREO DE LAS VARIANZAS SESGADA E INSESGADA Y DISTRIBUCION EN EL MUESTREO DE LA MEDIA

| *Muestra* (*Salarios de 4 trabajadores*) | *Media muestral* | *Varianza sesgada* $s^2 = \sum x^2/n$ | *Varianza insesgada* $\hat{s}^2 = \sum x^2/(n-1)$ |
|---|---|---|---|
| $1, 2, 3, 3 | $2.25 | 2.75/4 | 2.75/3 |
| 1, 2, 3, 4 | 2.50 | 5.00/4 | 5.00/3 |
| 1, 2, 3, 5 | 2.75 | 8.75/4 | 8.75/3 |
| 1, 2, 3, 4 | 2.50 | 5.00/4 | 5.00/3 |
| 1, 2, 3, 5 | 2.75 | 8.75/4 | 8.75/3 |
| 1, 2, 4, 5 | 3.00 | 10.00/4 | 10.00/3 |
| 1, 3, 3, 4 | 2.75 | 4.75/4 | 4.75/3 |
| 1, 3, 3, 5 | 3.00 | 8.00/4 | 8.00/3 |
| 1, 3, 4, 5 | 3.25 | 8.75/4 | 8.75/3 |
| 1, 3, 4, 5 | 3.25 | 8.75/4 | 8.75/3 |
| 2, 3, 3, 4 | 3.00 | 2.00/4 | 2.00/3 |
| 2, 3, 3, 5 | 3.25 | 4.75/4 | 4.75/3 |
| 2, 3, 4, 5 | 3.50 | 5.00/4 | 5.00/3 |
| 2, 3, 4, 5 | 3.50 | 5.00/4 | 5.00/3 |
| 3, 3, 4, 5 | 3.75 | 2.75/4 | 2.75/3 |
| Total, 15 muestras | 45.00 | 90.00/4 = 22.50 | 90.00/3 = 30.00 |
| Media de los estadísticos (valor esperado) | $\frac{45.00}{15} = \$3$ | $\frac{22.50}{15} = \$1.5$ | $\frac{30.00}{15} = \$2$ |
| Parámetro de la población | $\mu = \frac{18}{6} = \$3$ | $\sigma^2 = \frac{10}{6} = \$1.67$ | $\hat{\sigma}^2 = \frac{10}{5} = \$2$ |

## INTERVALOS DE CONFIANZA PARA $\mu$

El procedimiento para calcular el intervalo (a,b) que comprenda al parámetro con una probabilidad $1 - \alpha$ recibe el nombre de estimación por intervalos.

Probabilidad $( a < \mu < b ) = 1 - \alpha$

Probabilidad $( \overline{X} - Z_c * S / \sqrt{n} < \mu < \overline{X} + Z_c * S / \sqrt{n} ) = 1 - \alpha$

Probabilidad$(\underset{a}{\rule{3cm}{0.4pt}} < \mu < \underset{b}{\rule{3cm}{0.4pt}}) = 1 - \alpha$

a, b = límites del intervalo de confianza $\overline{X} \pm Z_c$ Error estándar

$Z_c$ = calificación estándar a partir de la cual el área vale $\alpha/2$

$\overline{X}$ = media muestral

S = desviación estándar

n = tamaño de muestra

b - a : es una medida de la precisión y

$1 - \alpha$ : es una medida de confiabilidad

**EJEMPLO**

$\overline{X} = 160$

$S^2 = 900$

$n = 36$

$$Pr(a < \mu < b) = 0.95$$

$$a = 160 - 1.96 * 30 / \sqrt{36} = 150.2$$

$$b = 160 + 1.96 * 30 / \sqrt{36} = 169.8$$

## ESTIMACION POR INTERVALOS DE LA DIFERENCIA DE DOS MEDIAS

Si $D = X_1 - X_2$ entonces $\overline{D} = \overline{X}_1 - \overline{X}_2$

y asumiendo independencia

$$S^2_{\overline{D}} = S^2_{\overline{X1}-\overline{X2}} = S^2_{\overline{X1}} + S^2_{\overline{X2}} = S^2_1 / n_1 + S^2_2 / n^2$$

El intervalo de confianza será

$$Pr\left(\overline{D} - Z_c \sqrt{S^2_1 / n_1 + S^2_2 / n^2} < ó < \overline{D} + Z_c \sqrt{S^2_1 / n_1 + S^2_2 / n^2}\right) = 1-\alpha$$

$$Pr\left(\underset{a}{\underline{\qquad\qquad}} < ó < \underset{b}{\underline{\qquad\qquad}}\right) = 1-\alpha$$

**EJEMPLO**

Continuando con el ejemplo de habichuela, en una estrategia de manejo de plagas diferente a la del agricultor, se obtienen los siguientes resultados:

| | n | $\overline{X}$ larvas/trifolio | $S^2$ |
|---|---|---|---|
| Agricultor | 12 | 22.24 | 935.944 |
| Alternativa | 12 | 3.9833 | 16.012 |

$$\overline{D} = \overline{\text{Agricultor}} - \overline{\text{Alternativa}} = 18.2567$$

$$S^2_{\overline{D}} = \frac{935.944}{12} + \frac{16.012}{12} = 79.3297$$

$$S_{\overline{D}} = 8.91$$

En consecuencia un intervalo de confianza del 95% para la diferencia es el siguiente:

$$Pr\ (\ 18.2567 - 1.96 * 8.91 < \delta < 18.2567 + 1.96 * 8.91\ ) = 0.95$$

$$Pr\ (\ 0.7931 < \delta < 35.72\ ) = 0.95$$

**ESTIMACION POR INTERVALOS DEL PRODUCTO/RAZON DE VARIABLES**

$$P = X * Y \qquad \overline{P} = \overline{X} * \overline{Y}$$

$$ES_p = ( \overline{Y^2}\, S^2_x / n_x + \overline{X^2}\, S^2_y / n_y + S^2_x\, S^2_y / n_x\, n_y )^{½}$$

$$R = X/Y \qquad \overline{R} = [\, \overline{X} / \overline{Y}\, ] + \underbrace{[\, \overline{X} / \overline{Y}\, ] * [\, S^2_y / Y^2 * n_y\, ]}_{\text{sesgo}}$$

$$ES_R = Y^{-2} \; ( S^2_x / n_x + R^2\, S^2_y / n_y )^{½}$$

Como en los casos anteriores

$$P ( a < P , R < b ) = 1 - \alpha$$

Cantidad : 12440.45 gramos / familia / mes ( Y )

$S^2$ : 158813281

Gasto : 1489.16 $ / familia / mes ( X )

$S^2$ : 2438937

número hogares : 429205

Se desea estimar un intervalo de confianza de 95% para el precio / kilo.

$$P = \frac{1489.16}{12.44045} + \frac{1489.16}{12.44045} * \frac{158.813281}{(12.44085)^2 \, 429205}$$

$= 119.69$ ( El valor correspondiente al sesgo es irrelevante)

ES(R) = 0.2663

$P_r\,( 119.69 \pm 1.96 * 0.2663 ) = 0.95$

## TRANSFORMACION DE VARIABLES

Las relaciones definidas anteriormente como combinación lineal de variables ( $y = a + bx$ ), Productos y Razones, constituyen la mayoría de las transformaciones requeridas para manejo habitual de datos:

- Deflactar precios

- Calcular precios relativos

- Homogenizar variables en términos de área, humedad, hogar, etc.

Las características básicas de las variables originales, como media y varianza pueden utilizarse en la forma descrita para encontrar los respectivos valores en la escala resultante, sin tener que volver a los valores individuales.

En igual forma pueden trabajarse los intervalos de confianza.

## NUMEROS INDICES

Cuando se dispone de una serie cronológica se puede simplificar la información, principalmente en términos de interpretación, recurriendo a los números índices ( usualmente porcentajes) referidos a un momento específico llamado " Período Base ".

- El valor del período base se toma como 100 cuando los datos se reducen a porcentajes.

- Expresan muy bien variaciones relativas en diferentes momentos o lugares. Representa la viariación promedio de un conjunto de valores en <u>dos</u> o más periodos diferentes.

- Hay índices <u>simples</u>, para una sola variable y <u>compuestos</u>, para varias. Estos pueden ser ponderados o no.

Indice Simple = 100 * Valor i / valor base

Ejemplo:
Precios deflactados

## ASPECTOS A TENER EN CUENTA

- Variables que deben incluirse para responder a los objetivos planteados.

- Los índices se construyen a partir de muestras, las cuales deben ser representativas de la población originaria. Es usual considerar la posibilidad de estratos.

- Para que el índice sea confiable, los datos deben ser comparables en términos de las variables que representan. Donde hay importantes cambios tecnológicos debe tenerse en cuenta para su interpretación ( Un computador de 1970 y otro en 1991 ). Crítico para índices de precios y de costos.

- El período base debe ser de estabilidad relativa y no demasiado antiguo.

- La ponderación determina la influencia de cada variable en el índice, por lo tanto debe considerarse su importancia o significación relativo para poder tener resultados con sentido.

## INDICES GLOBALES SIMPLES

$I_n \quad = \quad 100 * \Sigma P_n / \Sigma P_o$

$\Sigma P_o$ = Sumatoria de los valores en el periodo base

$\Sigma P_n$ = Sumatoria de los valores en el periodo n

Se supone que cada variable tiene la misma importancia.

## CARACTERISTICAS

- Fácil comprensión

- Supone igual importancia de cada producto

- Afectado por unidades de cada variable ( pesos ocultos )

**EJEMPLO :**

## INDICE GLOBAL SIMPLE

| Artículo | 1984 ( $P_o$ ) | 1987 ( $P_1$ ) | 1990 ( $P_2$ ) |
|---|---|---|---|
| X: $/tonelada | 1200 | 1500 | 1560 |
| Y: $/docena | 300 | 360 | 330 |
| Z: $/kilogramo | 500 | 600 | 570 |
| Gobal | 2000 | 2460 | 2460 |
| Indice | 100 | 123 | 123 |

## PROMEDIO SIMPLE DE PORCENTAJES RELATIVOS

- Se convierten los valores reales de cada variable en porcentajes con respecto al periodo base.

- Se promedian estos valores

$$P = 100 * \Sigma ( P_n / P_o ) / n$$

Para el caso anterior tendríamos:

| | $P_o / P_o$ | $P_1 / P_o$ | $P_2 / P_o$ |
|---|---|---|---|
| X | 100 | 125 | 130 |
| Y | 100 | 120 | 110 |
| Z | 100 | 120 | 114 |
| Global | 300 | 365 | 354 |
| Indice | 100 | 121.7 | 118 |

**CARACTERISTICAS**

- Insensible a las Unidades

- Supone que cada uno de los porcentajes relativos tiene igual importancia

## NUMEROS INDICES GLOBALES PONDERADOS

Implica la elección de pesos. Usualmente se hace así:

- Tome cantidades ( $q_o$ ) y precios ( $p_o$ ) en el período base

- Tome cantidades ( $q_n$ ) y precios ( $p_n$ ) en otros períodos

$P_l = 100 * \Sigma\, p_n\, q_o \,/\, \Sigma\, p_o\, q_o$

**INDICE DE LASPEYRES**

$P_p = 100 * \Sigma\, p_n\, q_n \,/\, \Sigma\, p_o\, q_n$

**INDICE DE PAASCHE**

El cálculo engorroso del denominador de $P_p$ lo hace poco usado.

## INDICE DE LASPEYRES

| | | | | | Valor de la cantidad $q_o$ a precios | | |
|---|---|---|---|---|---|---|---|
| | $q_o$ | $p_o$ | $p_1$ | $p_2$ | $p_o$ | $p_1$ | $p_2$ |
| X | 10 | 20 | 22 | 22 | 200 | 220 | 220 |
| Y | 20 | 10 | 12 | 15 | 200 | 240 | 300 |
| Z | 50 | 5 | 6 | 7 | 250 | 300 | 350 |
| Global | | | | | 650 | 760 | 870 |
| Indice | | | | | 100 | 118.5 | 133.8 |

**PROMEDIO PONDERADO DE PRECIOS RELATIVOS**

Haciendo extensión de los índices globales de Laspeyres y Paasche tendremos:

$$P_l = 100 * \Sigma ( P_n / P_o ) ( P_o q_o ) / \Sigma p_o q_o$$

$$P_p = 100 * \Sigma ( P_n / P_o ) ( P_o q_n ) / \Sigma P_o q_n$$

pero los resultados son idénticos.

## ALGUNAS CONSIDERACIONES SOBRE LA PRESENTACION DE DATOS

- **CONSIDERACIONES GENERALES**

- El nivel de detalle y desagregación depende del público al cual va dirigido el reporte oral o escrito.

- La definición de variables, los encabezamientos de las tablas y su configuración deben ser sencillos, pues el lector/auditorio no necesariamente dispone de conocimientos técnicos o matemáticos sofisticados.

- El lenguaje técnico debe ser explicado para lectores no profesionales.

## TABLAS - GRAFICAS

- El texto que acompañe las gráficas debe resumir algunos hechos importantes revelados por las cifras, indicando las conclusiones pertinentes.

- Gráficas y diagramas pueden ser usados para incrementar los niveles de comprensión.

- No debe haber redundancia entre el texto, las tablas y las gráficas.

- Los intervalos de clase de amplitud variable confunden al lector.

- Los límites de clase deben definirse sin traslape.

- El número de clases debe permitir claridad sin ser exageradamente breve o detallado. Debe reflejar la variabilidad de los datos.

- Las unidades de medida, el período o la zona representada por las cifras debe ser explícito, lo mismo que la fuente.

- Si una variable es categórica, para ella pueden hacerse análisis de frecuencias si es cardinal, pueden hacerse ademas otros análisis adicionales : medidas de tendencia central, dispersión, simetría, concentración, inferencias, etc.

Se debe dar a cada tipo de variable el trato justo. Es decir, no tratar como cardinal aquellas que no lo son.

- Si la presentación es en tabla cruzadas debe resaltarse que tipo de cifras se presentan:
  - frecuencias en la celda
  - promedios
  - porcentajes

- Si una de las dimensiones de la tabla es geográfica pueden formarse grupos por afinidad ( p.e: agrupar paises por continente ) si ello no oculta el propósito de la tabla ( p.e: jerarquizar una variable).

- Si la tabla es de frecuencia, deben incluirse los marginales, o sea total de cada fila y total de cada columna.

- Si el número de casos no permite una estimación confiable para alguna cifra este hecho debe indicarse con notas de pie de página.

## PORCENTAJES E INDICES

- Debe informase con respecto a que valor se calculan ( p. e: porcentajes por fila, por columna, con respecto al total o a un valor, columna o fila específico).