pacorabadan.com

4. Medidas de Dispersión

ESTADÍSTICA DESCRIPTIVA

DR. FRANCISCO RABADÁN PÉREZ

Índice

1. Medidas de Dispersión

2. Medidas de Dispersión Absolutas

3. Medidas de Dispersión Relativas

    3.1 Tipificación

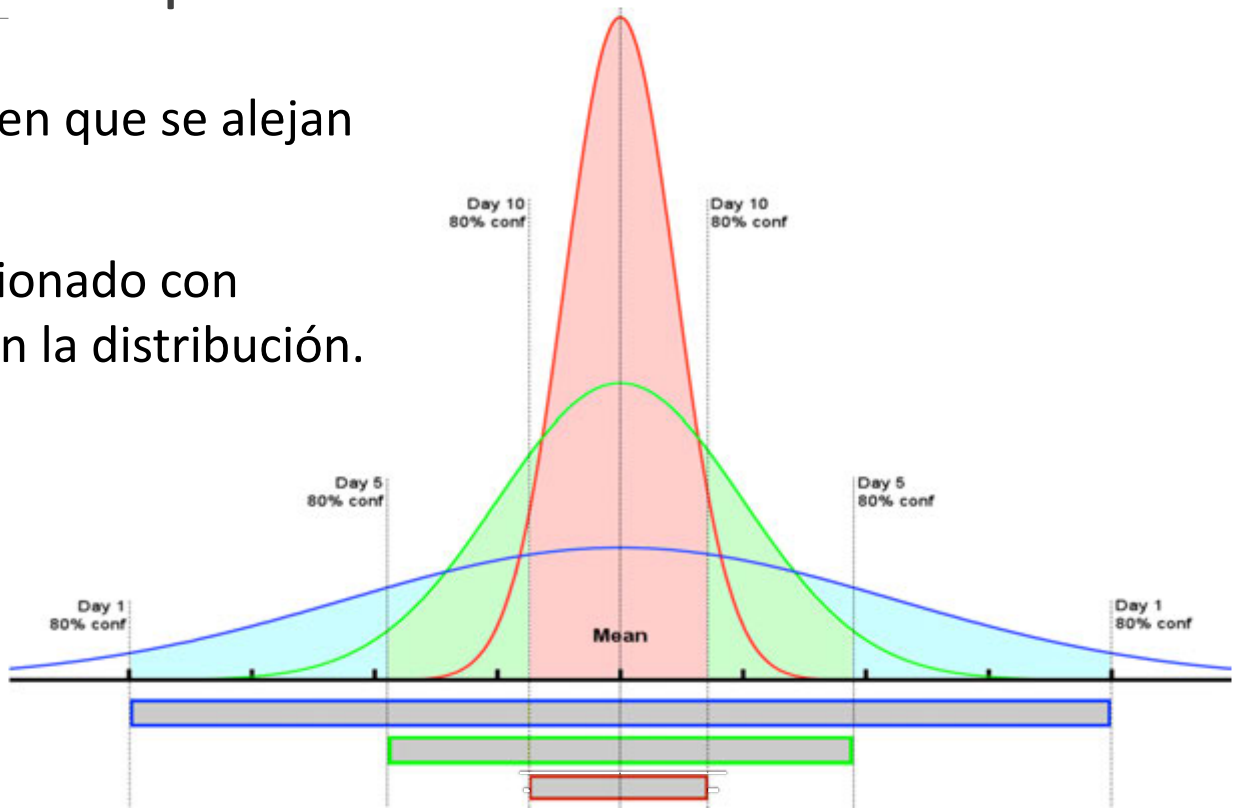
    3.2 Problemas al comparar dispersiones con la desviación estándar

    3.3 Coeficiente de Variación de Karl Pearson

1. Medidas de Dispersión

Objetivo: buscamos medir el grado en que se alejan entre sí los valores de la variable.

El concepto de dispersión está relacionado con separación de datos o variabilidad en la distribución.



Fuente: https://sites.google.com/site/estadisticaguileca/medidas-de-dispersion

1. Medidas de Dispersión

La representatividad de una medida de posición se mide como la separación (o dispersión) entre el conjunto de valores x_i y esa medida de posición.

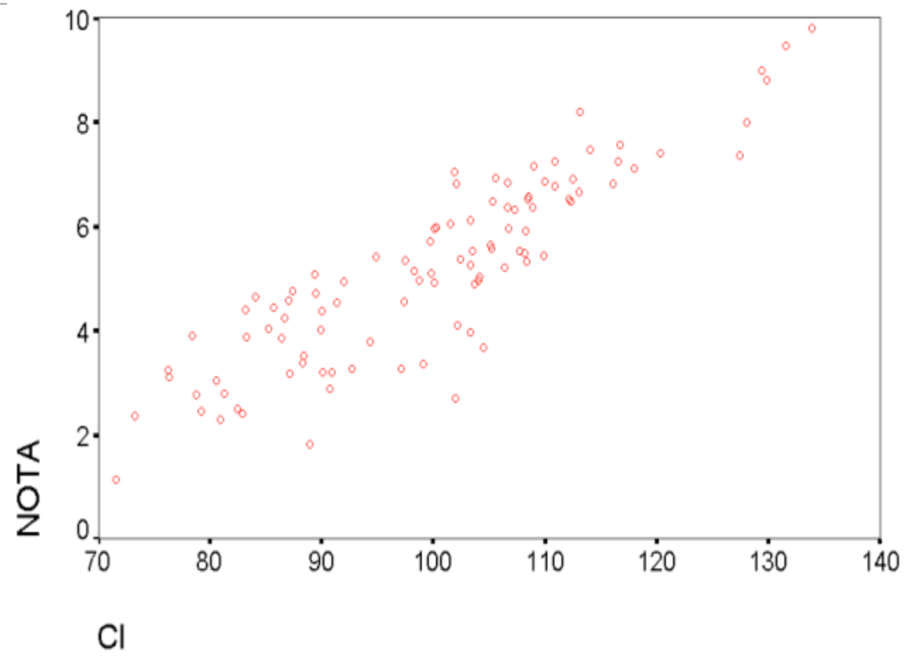


Diagrama de dispersión:
http://matematicas1bc.blogspot.com.es/2012/05/estadistica.html

1. Medidas de Dispersión

- **Medidas de Dispersión**
  - **Absolutas** (en unidades de x)
    - Recorrido (Re)
    - Recorrido Intercuartílico (RI)
    - Desviación media
      - Respecto a la media
      - Respecto a la mediana
    - Varianza
    - Desviación típica
  - **Relativas** (adimensionales)
    - Coeficiente de apertura
    - Recorrido semi-intercuartíclico
    - Coeficiente de variación de Pearson
    - Índice de dispersión respecto a la Mediana

2. Medidas de Dispersión Absolutas

Medida	Cálculo	Interpretación
Recorrido	$Re = x_n - x_1$	• Máxima diferencia que encontraremos entre dos valores cualesquiera de la distribución.
Recorrido Intercuartílico	$IQR = C_3 - C_1$	• Amplitud del intervalo en que está el 50% de los valores centrales
Desviación media respecto a la media	$D = \sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert \frac{n_i}{N}$	• Media aritmética de las distancias absolutas de los valores a la $\bar{x}$ • Recordemos $m_1=0$
Desviación media respecto a la mediana	$D_{Me} = \sum_{i=1}^{n} \lvert x_i - Me \rvert \frac{n_i}{N}$	• Media aritmética de las distancias absolutas de los valores a la Me

$\bar{x}$ hace mínimas las desviaciones cuadráticas pero la Me hace mínima las desviaciones absolutas, por tanto…

$$D_{Me} < D$$

2. Medidas de Dispersión Absolutas

Medida	Cálculo	Interpretación y consideraciones
Varianza	$S^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 \frac{n_i}{N}$	• Media de las desviaciones cuadráticas • **Difícil interpretación**: Unidades al cuadrado • No permite establecer comparaciones entre distintas magnitudes.
Cuasivarianza	$S_1^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 \frac{n_i}{N-1}$	• La veremos en mas profundidad en la Estadística II (Es mejor estimador que la varianza)
Desviación típica o estándar	$S = \left\lvert \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \frac{n_i}{N}} \right\rvert$	• Raíz cuadrada positiva de la varianza. • Se interpreta como lo que se alejan típicamente los valores x_i de $\bar{x}$ • **Fácil interpretación**: en las mismas unidades que la variable

K. Pearson la desarrollo en 1894 para solucionar el problema de que la media de las desviaciones respecto a $\bar{x}$ es cero (m_1=0).
Ver: http://www.expansion.com/diccionario-economico/desviacion-tipica.html

$$D_{Me} < D < S$$

(Martín Pliego, 2011; pág. 86)

2. Cálculo de la varianza (ejemplo)

Ejemplo:
Calculese la varianza de una distribución de frecuencias referente a los resultados obtenidos con 50 lanzamientos de un dado

x_i	n_i	$x_i n_i$	$x_i^2 n_i$
1	6	6	6
2	11	22	44
3	6	18	54
4	7	28	112
5	9	45	225
6	11	66	396
SUMAS	50	185	837

$$\boldsymbol{s^2 = a_2 - a_1^2} =$$

$$= \frac{\sum_{i=1}^{n} x_i^2 n_i}{N} - \left(\frac{\sum_{i=1}^{n} x_i n_i}{N} \right)^2 =$$

$$= \frac{837}{50} - \left(\frac{185}{50} \right)^2 = 3{,}05$$

(Martín Pliego, 2011; pág. 85)

2. Medidas de Dispersión Absolutas

Propiedades de la varianza:

- P-1: $S^2 \geq 0$
- P-2: S^2 es la **medida cuadrática de dispersión óptima**
- P-3: $S^2 = m_2 = a_2 - a_1^2 = a_2 - (\bar{x})^2$ (*"fórmula fácil de la varianza"*...)
- P-4: S^2 no se ve afectada por los cambios de origen
- P-5: S^2 queda afectada por cambios de escala de la siguiente forma

$$(x_i; n_i; S^2) \rightarrow (x_i' = x_i \cdot k, ; n_i; S'^2 = k^2 S^2)$$

S^2

varianza muestral

σ^2

varianza poblacional

2. Medidas de Dispersión Absolutas

Propiedades de la desviación típica:

- P-1: $S \geq 0$
- P-2: S es una medida de dispersión óptima
- P-3: $\mathrm{S} = \sqrt{a_2 - a_1^2} = \sqrt{a_2 - (\bar{x})^2}$
- P-4: S no se ve afectada por los cambios de origen
- P-5: queda afectada por cambios de escala: $S' = |k| \cdot S$

S

Desviación estándar muestral

σ^2

Desviación estándar poblacional

3. Medidas de Dispersión Relativas

Medida	Cálculo	Interpretación
Coeficiente de Apertura	$A = \frac{x_n}{x_1}$	• Presenta muchos inconvenientes (Vid. Martín-Pliego,211,pág. 87)
Recorrido Semi-Intercuartílico	$R_S = \frac{C_3 - C_1}{C_3 + C_1}$	• Pretende comparar la variación en el 50% central de la distribución con un teórico 100% de la distribución no afectado por valores extremos
Coeficiente de variación de Pearson	$CV = \frac{S}{\lvert\bar{x}\rvert}$	• Expresa la desviación típica en medias aritméticas. • Presenta problemas cuando $\bar{x} \simeq 0$ • Funciona muy bien cuando los rangos de las variables son similares.
Índice de dispersión respecto a la Me	$V_{Me} = \frac{D_{Me}}{Me} = \frac{\sum_{i=1}^{n} \lvert x_i - Me \rvert n_i}{N \cdot Me}$	• Mide la desviación media respecto a la mediana en medianas. • Es un concepto similar al de Pearson, pero presenta una mayor dificultad respecto a la generalización a la población.

3. Ejemplo

Para comparar los rendimientos entre empresas españolas y norteamericanas, pertenecientes a un sector muy especializado, se seleccionan 20 empresas con características semejantes de cada lugar, obteniéndose los siguientes resultados,

Empresas Españolas		Empresas Norteamericanas	
Beneficios (€) x_i	Frecuencias n_i	Beneficios ($) y_j	Frecuencias n_j
1.000.000	4	10.000	2
1.100.000	6	11.000	2
1.200.000	6	12.000	4
1.300.000	2	13.000	4
1.400.000	2	14.000	4
		15.000	2
		16.000	2

(Martín Pliego, 2011; pág. 89,90)

Se pide: obtener el rendimiento medio en cada país, precisando en cuál de los dos hay menos dispersión relativa.

a) en las empresas españolas

$$\bar{x} = \sum \frac{x_i n_i}{N} = 1.160.000$$

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2 n_i}{N} = 14.400.000.000$$

$$s_x = \sqrt{s_x^2} = 120.000$$

$$V_x = \frac{120.000}{1.1.60.000} = \frac{12}{16} = 0{,}1034$$

b) en las empresas norteamericanas

$$\bar{y} = \sum \frac{y_j n_i}{N} = 13.000$$

$$s_x^2 = \frac{\sum \left(y_j - \bar{y}\right)^2 n_i}{N} = 3.000.000$$

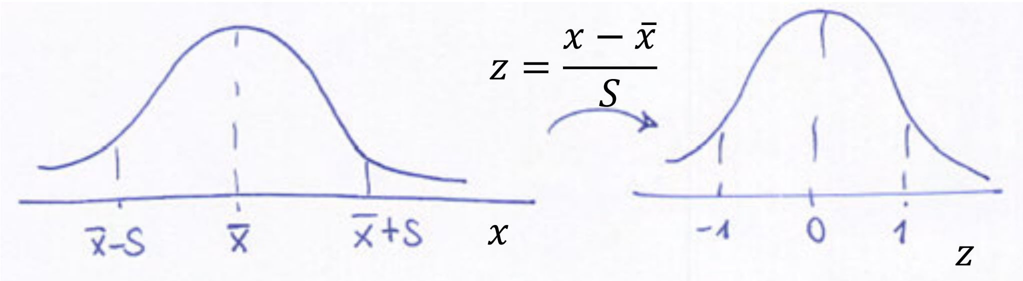
$$s_y = \sqrt{S_y^2} = 1.732{,}0508$$

$$V_y = \frac{1.732{,}0508}{13.000} = 0{,}1332$$

De la comparación entre los coeficientes de variación resulta que las empresas españolas $(0{,}1034 < 0{,}13)$ tienen menor dispersión relativa que las norteamericanas

3.1. Tipificación

IMPORTANTE



$$z = \frac{x - \bar{x}}{S}$$

Variable tipificada

- La tipificación transforma la variable tanto en escala como en origen.
- El valor z_i es equivalente a x_i pero en una distribución de media cero y desviación típica uno.
- Es especialmente útil para comparar la posición que ocuparía un individuo en diferentes colectivos ; teóricamente, que nota hubiera sacado un alumno si hubiera estado en un grupo distinto….

CUIDADO: habitualmente se confunde la variable Z con la N(0,1), pero la tipificación sólo afecta al origen y a la escala, no a la forma de la distribución.

3.1. Tipificación

IMPORTANTE

Comparando valores de distintas variables de forma adimensional con la tipificación

$$z = \frac{x - \bar{x}}{S}$$

Variable tipificada

Grupo Tomás	Grupo Gaspar
$\bar{x} = 5{,}5$	$\bar{y} = 4$
$S_x = 1$	$S_y = 4$
Nota: 4	Nota: 4,5

Ejemplo: Queremos comparar de forma relativa la nota que dos alumnos han sacado en diferentes grupos para determinar cual ocupa una posición relativa mejor.

$$z_{Tomás} = \frac{x - \bar{x}}{s_x} = \frac{4 - 5{,}5}{1} = -1{,}5$$

$$z_{Gaspar} = \frac{y - \bar{y}}{s_y} = \frac{4{,}5 - 4}{4} = 0{,}125$$

Gaspar ocupa una posición que supera levemente a la media, mientras que Tomás ocupa una posición inferior a la media alejándose negativamente en 1,5 desviaciones típicas en una distribución de media cero y desviación típica 1.

Conclusión: **Gaspar es mejor alumno que Tomás. Si hubieran estado en un grupo similar, le superaría en 1,625 desviaciones típicas.**

3.1. Tipificación

IMPORTANTE

Traslación equivalente a otras distribuciones basándonos en la tipificación

Grupo Tomás	Grupo Gaspar
$\bar{x} = 5{,}5$	$\bar{y} = 4$
$S_x = 1$	$S_y = 4$
Nota: 4	Nota: 4,5

$$z_{Tomás} = \frac{x - \bar{x}}{s_x} = \frac{4 - 5{,}5}{1} = -1{,}5$$

$$z_{Gaspar} = \frac{y - \bar{y}}{s_y} = \frac{4{,}5 - 4}{4} = 0{,}125$$

$$x = zS + \bar{x}$$

Calcular x en base a la variable tipificada

Si quisiéramos encontrar la nota de Tomás en el grupo de Gaspar

$$y_{tomás} = -1{,}5 * 4 + 4 = -2$$

Si quisiéramos encontrar la nota de Gaspar en el grupo de Tomás

$$x_{Gaspar} = 0{,}125 * 1 + 5{,}5 = 5{,}625$$

3.2. Problemas al comparar dispersiones con la desviación estándar

S

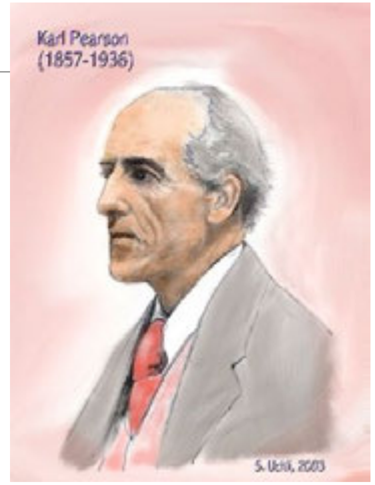
no nos sirve

Si dos variables vienen expresadas en distintas unidades de medida

(no se pueden comparar años y cm).

si los promedios son distintos.

Necesitamos una medida de dispersión adimensional, es decir, que no se vea afectada por las unidades de medida.



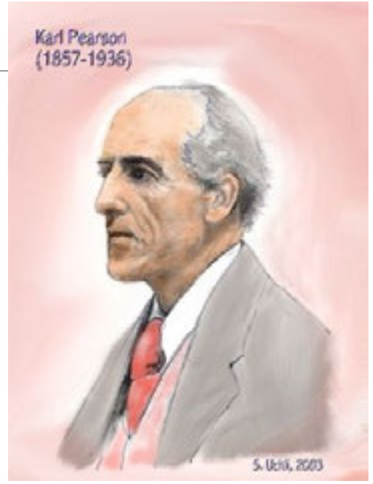
$$CV = \frac{S}{|\bar{x}|}$$

Coeficiente de variación de Pearson

3.2. Aplicación del Coeficiente de Variación de Karl Pearson.

IMPORTANTE

- **Características**
  - Número de veces que s contiene a $\bar{x}$
  - Si aumenta el CV:
    - aumenta la dispersión
    - $\bar{x}$ es menos representativa
    - CV>0 (cuidado si $\bar{x} \simeq 0$)
- **Ventajas**
  - Adimensional
  - Intervienen todos los valores y sus frecuencias
- **Inconvenientes**
  - Si $\bar{x}$=0 (habitual en la variable tipificada)
  - No es invariante frente a cambios de origen
  - ¿Frente a cambios de escala?



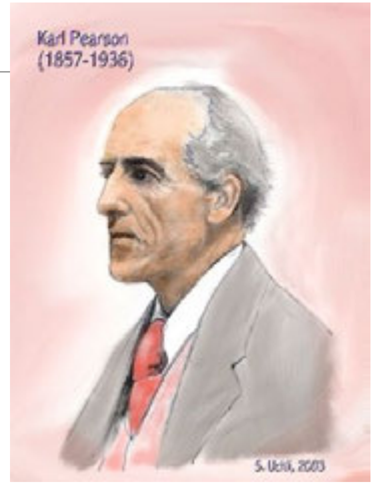
$$CV = \frac{S}{|\bar{x}|}$$

Coeficiente de variación de Pearson

IMPORTANTE

3.2. Interpretación del CV

	$\bar{x}$	S^2
$0<CV<{}^{1}/_{3}$	$\bar{x}$ muy representativa	Poca dispersión
${}^{1}/_{3}<CV<{}^{3}/_{4}$	$\bar{x}$ medianamente representativa	Dispersión Media
$CV>{}^{3}/_{4}$	$\bar{x}$ poco representativa	Mucha dispersión



$$CV = \frac{S}{|\bar{x}|}$$

Coeficiente de variación de Pearson

Textos recomendados

- Martín-Pliego, *Introducción a la Estadística Económica y Empresarial*, Editorial AC, 2011, 3ª Edición

Ejercicios y prácticas

Ejercicios del 1 al 5 (Martín-Pliego, 2011; pág. 90-97)

Ejercicios en clase

Recursos web y aula virtual.