Tema 1: Estadística descriptiva

Razonamiento estadístico

Analizar datos que se pueden utilizar de forma engañosa.

Ejemplo: el 60% de los ingenieros que usan gafas, son informáticos.

Ing. Gafas		Ing. No Gafas	
60% Infor.	40% Infor.	90% Infor.	10% Infor.

Ingenieros

Como hay más ingenieros informáticos que otro tipo, es normal que más lleven gafas

Estado descriptivo

* Tipos de variables: nominal, ordinal (cualitativas).
  discreta, continua (cuantitativas).

* Tabla Estadística:

- X : Variable numérica (más adelante habrá que decir si es discreta).
- n_i: Frecuencia absoluta.
- f_i: Frecuencia relativa. $f_i = \frac{n_i}{N}$
- N: Tamaño muestral (cantidad total de datos). $\sum_i n_i = N$

$$\sum_i f_i = \sum_i \frac{n_i}{N} = \frac{1}{N} \sum_i n_i = \frac{N}{N} = 1$$

- N_i : Frecuencia absoluta acumulada. $N_i = \sum_{j=1} n_j$
- F_i: Frecuencia relativa acumulada. $F_i = \frac{N_i}{N}$

Ejemplo: la siguiente tabla representa el nº de archivos que se mandaron a una impresora en 30 días consecutivos.

X = nº archivos

x_i	n_i	f_i	N_i	F_i
20 - 30	4	4/50 0'08	$n_1 = N_1$ 4	0'08
30 - 40	7	7/50 0'14	4+7 11	0'08+0'14 0'22
40 - 50	10	10/50 0'2	11+10 21	0'22+0'2 0'42
50 - 60	9	0'18	21+9 30	0'42+0'18 0'60
60 - 70	8	8/50 0'16	30+8 38	0'60+0'16 0'76
70 - 80	7	0'14	38+7 45	0'76+0'14 0'90
80 - 90	5	5/50 0'1	45+5 50	0'9+0'1 1

$\sum n_i = 30$ $\sum f_i = 1$ Casilla = N Casilla = 1

N = 50

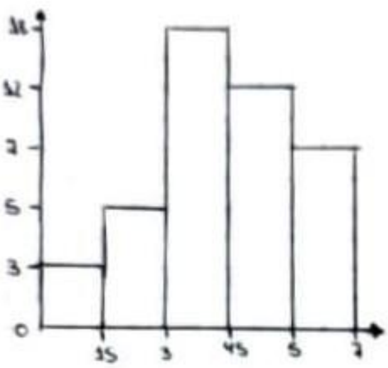
- Histograma

- a_i: amplitud de intervalo.
- f_i: $\propto$ área rectángulo (a_i, h_i). $f_i = k \cdot a_i \cdot h_i$

  $\sum_i f_i = 1 = k(4'5 + 7'5 + 27 + 6 + 14) = k\,59 \rightarrow k = 1/59$
- h_i: altura del histograma. $h_i = \dfrac{f_i}{\underset{k=1}{k} \cdot a_i}$ → Para pasar de tabla a histograma

Vamos a estudiar algunos parámetros (o estadísticos) de variables que me proporcionan información sobre los datos.

Ejemplo: lo pasamos a tabla estadística



→

I_i	h_i	a_i	f_i	f
$[0, 1'5]$	3	1'5	$k \cdot 3 \cdot 1'5$	0'076
$(1'5, 3]$	5	3 - 1'5	$k \cdot 5 \cdot 1'5$	0'127
$(3, 4'5]$	18	4'5 - 3	$k \cdot 18 \cdot 1'5$	0'457
$(4'5, 5]$	12	5 - 4'5	$k \cdot 12 \cdot 0'5$	0'101
$(5, 7]$	7	7 - 5	$k \cdot 7 \cdot 2$	0'237

PREGUNTAS IMPORTANTES

1. Dado un conjunto de datos $\{d_1, d_2 \dots d_n\}$ y fijado un dato (d_i), ¿qué posición tiene ese dato (d_i) en esa lista (L)?
   Para eso están las medidas de posición (parámetros o estadísticos).

2. Dado un conjunto de datos, ¿cómo de concentrados o dispersos están?
   Para responder a esto se usan las medidas de dispersión.

3. ¿Qué forma tienen los datos?
   A esto responden las medidas de forma.

MEDIDAS DE POSICIÓN

• Moda: dado un conjunto de datos, la moda es el dato que se repite más veces.

- Observaciones sobre la moda:

• No tiene por qué ser un único valor. Si se da el caso de que hay 2 valores que se repiten el mismo nº de veces, ambos serán la moda.

• La moda puede no existir. Si todos los valores se repiten el mismo nº de veces.

x_i	n_i
2	6
3	6
4	3

$M_o = \{2, 3\}$

x_i	n_i
2	6
3	6
4	6

No hay moda

• Media (aritmética): "es el parámetro más representativo de un conjunto de datos".

• Dado una lista de datos $\{d_1, d_2 \dots d_n\}$, la media $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} d_i$.

• Dado una tabla estadística, la media $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \cdot n_i$ o $\bar{x} = \sum_{i=1} x_i \cdot f_i$

- Observaciones sobre la media (aritmética):

• La media no tiene por qué coincidir con ninguna de las modalidades de la variable. $7 \mid 8 \rightarrow (7+8)/2 = 7'5 \rightarrow$ no coincide con 7 ni 8.

• $\min_j \{x_j\} \leq \bar{x} \leq \max_j \{x_j\} \rightarrow$ Tiene que estar entre el máx y el mín.

Demostración:

$\bar{x} = \frac{1}{N} \sum_{i=1} x_i n_i = \frac{1}{N} (x_1 n_1 + x_2 n_2 + \dots + x_n n_n)$ $\quad x^* = \max_{j=1} \{x_j\}$

Como $x_1 \leq x^*$ puedo decir que: $\frac{1}{N} (x_1 n_1 + x_2 n_2 + \dots) \leq \frac{1}{N} (x^* n_1 + x_2 n_2 + \dots) \leq \dots$

Por esto $x^* = \max_{j=1} \{x_j\} = x^* \frac{1}{N} (n_1 + n_2 + \dots)$

$\bar{x} = \frac{1}{N} \sum_{i=1} x_i n_i \leq \frac{1}{N} \sum_{i=1} x^* n_i = \frac{1}{N} x^* \sum_{i=1} n_i = x^* \underbrace{\sum_{i=1} \frac{n_i}{N}}_{1} = x^* = \max_{j=1} \{x_j\}$

• La media es un parámetro poco robusto (muy sensible a datos anómalos o posibles errores). → Outliers (picos en las gráficas). Estos picos no deberíamos tenerlos en cuenta en las medias.

x_i	n_i
1€	1
2€	3
3€	2

$\bar{x} = (1 \cdot 1 + 2 \cdot 3 + 3 \cdot 2)/6 = 2'16 \dots$

x_i	n_i
10€	1
20€	3
30€	2

$\bar{x} = (1 \cdot 1 + 20 \cdot 3 + 3 \cdot 2)/6 = 11'16 \dots$

Por esto hay que filtrar estos datos anómalos.

* Mediana: es un parámetro que (se encuentra en medio de los datos) deja la misma cantidad de datos a la izquierda y a la derecha.

- Observaciones sobre la mediana:
  - Si el nº de datos es par cogemos los 2 valores de en medio y hacemos la media.
    $\{-3, -3, 0, 0, 0, \underline{0, 1}, 1, 1, 2, 2, 3\} \rightarrow (0+1)/2 \rightarrow Me = 0'5$
  - Si los datos con los que trabajo son cualidades, se haría del mismo modo si fuesen impares.
    {Rubio, <u>Moreno</u>, Castaño} → Me = Moreno
  - Si son pares, se hará la media a los valores RGB o se le asignaran nºs del 0 al x a los que luego se les hará la media.
  - La mediana no tiene por qué coincidir con los datos.
  - La mediana es menos sensible (robusta) a los outliers que la m

- Algoritmo para calcular la mediana:
  - Si me dan una lista de datos $\{d_1, d_2, \ldots d_n\}$
    1. Ordena la lista de menor a mayor.
    2. Cuento cuantos datos tiene (N), es decir, el tamaño muestral.
    3. Si N es par: $Me = (d_k + d_{k+1})/2$, donde $k = N/2$.
       Si N es impar: $Me = d_k$, donde $k = \left[N/2\right] + 1$.
       Parte entera de la división
  - Si me dan una tabla estadística
    - Paso la tabla a lista (cuando es corta).
    - 1. Ordeno la tabla.
      2. Averiguo el tamaño muestral (N).
      3. N/2 para averiguar k.
      4. Nos fijamos en la frecuencia acumulada y hallamos la mediana como anteriormente.

Ejemplo:

x_i	n_i	N_i
-3	2	2
0	1	3
1	3	6

$\frac{N}{2} = \frac{6}{2} = 3 = k$

$Me = \frac{d_3 + d_4}{2} = \frac{0+1}{2} = 0'5$

$\{-3, -3, \underline{0, 1}, 1, 1\} \rightarrow Me = (0+1)/2 = 0'5$

Ejemplo completo: Dado el siguiente conjunto de datos, resultado de una variable: {1, 1, 0, -3, 1, 0, 2, 2, -3, 0, 0}

a) Construye una tabla estadística

x_i	n_i	f_i
-3	2	0'18
0	4	0'36
1	3	0'27
2	2	0'18

$N = 11$

b) Calcula la moda.

$M_o = 0$, el valor que más se repite

c) Calcula la media.

$\bar{x} = \frac{1}{11}(-3\cdot 2 + 0\cdot 4 + 1\cdot 3 + 2\cdot 2) = \frac{1}{11}$

d) Calcula la mediana.

$\{-3, -3, 0, 0, 0, \underline{0}, 1, 1, 1, 2, 2\} \rightarrow M_e = 0$

* El concepto de mediana puede generalizarse dando lugar a:

- **Cuantil**: dado un valor entre 0 y 1 $[C \in (0,1)]$, el cuantil es aquel valor que deja a su izquierda una proporción C y a su derecha una proporción 1-C. Si $C = 0'5 \rightarrow$ Cuantil $(0'5) = M_e$

  - **Algoritmo cuantil**:

    1. Ordeno la lista (N elementos)

    2. Multiplico $C \cdot N$ y el resultado lo descompongo como parte entera más parte decimal. $C \cdot N = E + D$

       $N = 11$, $C = 0'3$; $0'3 \cdot 11 = 3'3 = 3 + 0'3$

    3. Si $D = 0$ (No hay decimales), $C = (d_E + d_{E+1})/2$

       Si $D \neq 0$ (Hay decimales), $C = d_{E+1}$

- **Cuartil**: 1^er^ cuartil: $Q_1 =$ Cuantil $(0'25)$

  2º cuartil: $Q_2 =$ Cuantil $(0'5) = M_e$

  3^er^ cuartil: $Q_3 =$ Cuantil $(0'75)$

- **Decil (k)**: $D_k =$ Cuantil $(^k/_{10})$

- **Percentil (k)**: $P_k =$ Cuantil $(^k/_{100})$

Ejercicio pag 41 (diapositivas)

Calcular: P_{10}, Q_2, D_7

$P_{10} = C\left(\frac{10}{100}\right) \Rightarrow 0'1 \cdot 50 = 5 \rightarrow P_{10} = C(0'1) = \frac{d_5 + d_6}{2} = \frac{18+20}{2} = 19$

$Q_2 = C(0'5) \Rightarrow 0'5 \cdot 50 \Rightarrow M_e = 63'5 = Q_2$

$D_7 = C\left(\frac{7}{10}\right) = C(0'7) \Rightarrow 0'7 \cdot 50 \rightarrow D_7 = \frac{d_{35} + d_{36}}{2} = \frac{80+81}{2} = 80'5$

$\rightarrow$ Supongamos que se añade $d_{51} = 100$. Calcula D_7

$D_7 = C\left(\frac{7}{10}\right) = C(0'7) \Rightarrow 0'7 \cdot 51 \Rightarrow D_7 = d_{35+1} = 81$

MEDIDAS DE DISPERSIÓN

* Rango (o intervalo): llamamos rango o intervalo de un conjunto de datos a la diferencia entre el mayor y el menor.

Ejemplo: $\{0, 1, 7\}$; Rango $= 7 - 0 = 7$
$\{-3, 0, 7\}$; Rango $= 7 - (-3) = 10$

- Observaciones sobre el rango:
  - A mayor rango, mayor dispersión.
  - Si Rango $= 0 \rightarrow$ todos los datos coinciden.
  - Es un estadístico poco robusto frente a outliers.

Para arreglar ese problema, se definen:
- Rango intercuartílico: $RIC = Q_3 - Q_1$
- Rango intercentílico: $P_{99} - P_1$ (percentiles)

Ambos estadísticos son más robustos a outliers aunque puede hacerse algo mejor.

* Método para clasificar outliers: dado un conjunto de datos $D = \{d_1, d_2, \dots d_n\}$.

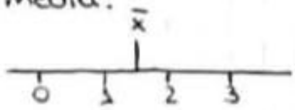
$$\text{Outliers} = \{d \in D \mid d < Q_1 - 3(Q_3 - Q_1)\} \cup \{d \in D \mid d > Q_3 + 3(Q_3 - Q_1)\}$$
$$D \cap [[Q_3 + 3(Q_3 - Q_1), Q_1 - 3(Q_3 - Q_1)]]$$

Ejemplo: $D = \{-10, 1, 2, 3, 10\}$ Filtra los valores anómalos usando el criterio anterior.

$$\left.\begin{array}{l} Q_1 = 5 \cdot 0'25 = 1'25 \rightarrow 1 \\ Q_2 = 5 \cdot 0'5 = 2'5 \rightarrow 2 \\ Q_3 = 5 \cdot 0'75 = 3'75 \rightarrow 3 \end{array}\right\} \begin{array}{l} [[Q_3 + 3(Q_3 - Q_1), Q_1 - 3(Q_3 - Q_1)]] \\ [[3 + 3(3-1), 1 - 3(3-1)]] = (9, -5) \end{array}$$

Por tanto los valores anómalos son $\{-10, 10\}$.

Volviendo a la dispersión, otra forma de ver si un conjunto es disperso o no podría ser viendo cómo de lejos se encuentran los datos con respecto al centro/media.

Ejemplo: 

$\bar{x} = 1'5$

$$\left.\begin{array}{l} \cdot \text{ Distancia de "3" a } \bar{x} = |3 - 1'5| \\ \cdot \text{ Distancia de "2" a } \bar{x} = |2 - 1'5| \\ \cdot \text{ Distancia de "1" a } \bar{x} = |1 - 1'5| \\ \cdot \text{ Distancia de "0" a } \bar{x} = |0 - 1'5| \end{array}\right\} \frac{|3-1'5| + |2-1'5| + |1-1'5| + |0-1'5|}{4} = 1$$

* Desviación media: $Dm = \frac{1}{N}\sum_{i=1}^{N} |di - \bar{x}|$; $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} di$

- Si vienen como tabla estadística: $Dm = \frac{1}{N}\sum_{i=1}^{N} |x_i - \bar{x}| \cdot n_i$; $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \cdot n_i$

- Desviación media con respecto a un valor p: $Dm(p) = \frac{1}{N}\sum_{i=1} |x_i - p| \cdot n_i$

- Observaciones sobre la desviación media:

$$\bullet \begin{cases} DM: \mathbb{R} \rightarrow \mathbb{R} \\ \quad p \rightarrow DM(p) = \frac{1}{N}\sum_{i=1} |x_i - p| \cdot n_i \end{cases}$$

* Error cuadrático medio respecto de un valor p: $ECM = \frac{1}{N}\sum_{i=1} (x_i - p)^2 \cdot n_i$

* Varianza: $ECM(\bar{x}) = \frac{1}{N}\sum (x_i - \bar{x})^2 \cdot n_i = V(\bar{x})$

* Desviación típica: $\sigma_{\bar{x}} = \sqrt{V(\bar{x})}$

Ejercicio 1: Descomposición de la varianza

$$V(\bar{x}) = \sum_{i=1} (x_i - \bar{x})^2 \cdot f_i = \dots = \left(\sum_{i=1} x_i^2 \cdot f_i\right) \cdot (\bar{x})^2$$

$$\sum_{i=1} (x_i - \bar{x})^2 f_i = \sum_{i=1} (x^2 + \bar{x}^2 - 2\bar{x}x_i) \cdot f_i = \sum x_i^2 \cdot f_i + \sum \bar{x}^2 \cdot f_i - \sum 2\bar{x}x_i \cdot f_i =$$

$$= \sum x_i^2 \cdot f_i + \bar{x}^2 \cdot \underbrace{\sum f_i}_{1} - \underbrace{2\bar{x}\sum x_i f_i}_{-2\bar{x}^2} = \left(\sum x_i^2 f_i\right) - (\bar{x}^2) = Var(x^2) = \sum x_i^2 \cdot f_i - \bar{x}^2$$

Ejercicio 2

$$\left.\begin{array}{l} V(\bar{x}) = 0 \\ x_i = cte \end{array}\right| \; V(\bar{x}) = 0 = \sum_{i=1} (x_i - \bar{x})^2 \cdot f_i = 0 \rightarrow x_i - \bar{x} = 0 \;\; \forall i = 1, 2, \dots$$

$$x_i = \bar{x} \;\; \forall i$$

Ejercicio 3: Linealidad de la media de una variable

$\bar{x} \sim$

x_i	n_i
x_1	n_1

, define $y = ax + b \longrightarrow$

y_i	n_i
$y_1 = ax_1 + b$	n_1

$$\bar{y} = a\bar{x} + b \rightarrow \bar{y} = \sum y_i f_i = \sum (ax_i + b) f_i = \sum a x_i f_i + \sum b f_i = a \underbrace{\sum x_i f_i}_{\bar{x}} + b \underbrace{\sum f_i}_{1} =$$

$$= a\bar{x} + b$$

Ejercicio 4

$V(\bar{x})$ ¿$V(\bar{y})$?

$$\frac{1}{N}\sum (x_i - \bar{x})^2 n_i \qquad \sum (y_i - \bar{y})^2 f_i = \sum ((ax_i + b) - \underbrace{(a\bar{x} + b)}_{\bar{y}})^2 f_i = \sum \left[a(x_i - \bar{x})\right]^2 f_i =$$

$$= \sum a^2 (x_i - \bar{x})^2 f_i = a^2 \sum (x_i - \bar{x})^2 f_i = a^2 V(\bar{x}) \Rightarrow V(\bar{y}) = a^2 V(\bar{x})$$

* Comparar dos muestras distintas

Quiero ver cual de las muestras, A y B, tiene más dispersión. Se usa:

* Coeficiente de variación de Pearson:

$$CV_A = \frac{\sigma_A}{\bar{A}} \ (\cdot 100\%) \ , \quad CV_B = \frac{\sigma_B}{\bar{B}}$$

Ejercicio página 52 (diapositivas)

$$\begin{cases} \bar{A} = 1495 \text{ horas} & \sigma_A = 280 \text{ horas} \\ \bar{B} = 1875 \text{ horas} & \sigma_B = 310 \text{ horas} \end{cases}$$

$$\left.\begin{array}{l} CV_A = \frac{280}{1495} \cdot 100 = 18'73\% \\ CV_B = \frac{310}{1875} \cdot 100 = 10'53\% \end{array}\right| CV_A > CV_B \text{ , A es más disperso, menos fiable.}$$

* Comparar dos individuos de dos muestras distintas

Una media que se usa para este tipo de comparación es la tipificación.

* Tipificación: la tipificación de la variable x es otra variable z.

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

Ejemplos:

* La nota media de una clase $\bar{A} = 7$ y $\sigma_A = 1$; $\bar{B} = 7$ y $\sigma_B = 0'5$. Se coge un alumno del grupo A con nota $7'5$; $a = 7'5$. Alumno grupo B; $b = 7'3$. ¿Quién ha ofrecido un mejor rendimiento?

Variable A → $a_i \mid n_i$

↓ tipificación

$$Z_A = \frac{A-7}{1}$$

nota

$$\frac{7'5-7}{1} = 0'5 \rightarrow 7'5 = 0'5 \cdot 1 + 7$$

desviación atípica

0'5 veces por encima de la media

Variable B → $b_i \mid n_i$

$$Z_B = \frac{B-7}{0'5}$$

$$\frac{7'3-7}{0'5} = 0'6 \rightarrow 7'3 = 7 + 0'6 \cdot 0'5$$

0'6 veces por encima de la media

* Una marca de relojes A dada con $\begin{cases} \bar{A} = 7\mu s;\ \sigma_A = 1\mu s,\ a = \text{error } 7'5\mu s. \\ \bar{B} = 7\mu s;\ \sigma_B = 0'5\mu s,\ b = \text{error } 7\mu s. \end{cases}$

$$Z_A = \frac{A-7}{1} = \frac{7'5-7}{1} = 0'5 \rightarrow 7'5 = 7 + 0'5 \cdot 1$$

$$Z_B = \frac{B-7}{0'5} = \frac{7-7}{0'5} = 0 \rightarrow 7 = 7 + 0 \cdot 0'5 \rightarrow$$ 0 veces por encima de la media por lo que es mejor.

- Momento central de orden r

$$\mu_r = \sum_{i=1} (x_i - \bar{x})^r \cdot f_i$$

- Observaciones sobre momento central
  - $\mu_0 = 1$
  - $\mu_1 = \sum_{i=1} (x_i - \bar{x}) f_i = \sum x f_i - \sum \bar{x} f_i = 0$
  - $\mu_2 = \sum_{i=1} (x_i - \bar{x})^2 \cdot f_i =$ Varianza

- Momento ordinario de orden r

$$m_r = \sum_{i=1} x_i^r \cdot f_i$$

- Observaciones sobre el momento ordinario
  - $m_0 = 1$
  - $m_1 = \bar{x}$
  - $m_2 = \sum_{i=1} x_i^2 \cdot f_i$

$Var(x) = \sum x_i^2 f_i - (\bar{x})^2$

$\mu^2 = m_2 - m_1^2$

$\mu_2 = \sum_{i=1} (x_i - \bar{x})^2 \cdot \frac{1}{N}$

- $\mu_3 = m_3 - 3m_2\bar{x} + 2\bar{x}^3$
- $\mu_4 = m_4 - 4m_3\bar{x} + 6m_2\bar{x}^2 - 3\bar{x}^4$

MEDIDAS DE FORMA

Una distribución de datos puede estar:

- Sesgada a la derecha. 
- Sesgada a la izquierda. 
- Simétrica 

Una distribución de datos tiene un apuntamiento o curtosis

- Platicúrtica 
- Mesocúrtica 
- Leptocúrtica 

Los criterios para determinar la forma de una distribución son:

- Criterio de simetría:

- Coeficiente de asimetría de Pearson: $A_p = \frac{\bar{x} - M_o}{\sigma}$
  - $A_p > 0 \rightarrow$ Simetría a la derecha / Cola a la derecha
  - $A_p < 0 \rightarrow$ Simetría a la izquierda / Cola a la izquierda
  - $A_p = 0 \rightarrow$ Simétrica

- Coeficiente de asimetría de Fisher : $g_1 = \frac{\mu_3}{\sigma_3}$
  - $g_1 > 0 \rightarrow$ Cola a la derecha
  - $g_1 = 0 \rightarrow$ Simétrica

- Criterio de apuntamiento o curtosis:
  - Coeficiente de aplastamiento de Fisher : $g_2 = \frac{\mu_4}{\sigma_4} - 3$
  - $g_2 > 0 \rightarrow$ Leptocúrtica
  - $g_2 = 0 \rightarrow$ Mesocúrtica
  - $g_2 < 0 \rightarrow$ Platicúrtica

* Media ponderada:

x_i	n_i	w_i

$\rightarrow$ pesos

$$\text{Media ponderada} = \frac{\sum x_i \frac{n_i}{N} \cdot w_i}{\sum w_i}$$

- Otras medias:
  - Aritmética : $\bar{x} = \sum x_i \frac{1}{n_i}$
  - Geométrica $= (x_1^{n_1} \cdot x_2^{n_2} \cdot \ldots x_k^{n_k})^{1/N}$; $N = n_1 + \ldots + n_k$
  - Cuadrática $= \sqrt{\sum x_i^2 \cdot \frac{n_i}{N}}$ $\rightarrow$ Para trabajar con errores

Ejemplos:

$$\left.\begin{array}{llll} * \ 1^{er} \text{ año} & \text{aumenta} & \text{en} & 10\% \\ 2^{o} \text{ año} & \text{"} & \text{en} & 20\% \\ 3^{er} \text{ año} & \text{"} & \text{en} & 30\% \\ 4^{o} \text{ año} & \text{"} & \text{en} & 40\% \end{array}\right\} N = 50 \left\{\begin{array}{l} 1^{er} \text{ año} = 1'1 \cdot 50 = 55 \\ 2^{o} \text{ año} = 1'2 \cdot 55 = 66 \\ 3^{er} \text{ año} = 1'3 \cdot 66 = 85'8 \\ 4^{o} \text{ año} = 1'4 \cdot 85'8 = 119'32 \end{array}\right.$$

* $\frac{10 + 20 + 30 + 40}{4} = 25$, Sabiendo solo eso, ¿cuál es la población a los 4 años? $(1'25)^4 \cdot 50 \simeq 122$ personas

* $\sqrt[4]{(1'1) \cdot (1'2) \cdot (1'3) \cdot (1'4)} \simeq 1'24$

$(1'24)^4 \cdot 50 \simeq 120$ personas

Los 3 ejemplos dan aprox. 120 personas en el cuarto año.

- Media armónica : $H = \frac{N}{\sum_{i=1} \frac{n_i}{x_i}}$

$$\left.\begin{array}{l} 1h \rightarrow 60 \text{ Km/h} \\ 1h \rightarrow 20 \text{ Km/h} \end{array}\right\} 2h \text{ recorre } 80 \text{ Km} \rightarrow V_{media} = \frac{80}{2} = 40 \text{ Km/h}$$