STATISTICA DESCRITTIVA

SIMBOLOGIA

fa_i	= frequenza assoluta
f_i	= frequenza relativa
$\overline{x}_i$	= valori argomentali della variabile
N	= numerosità del campione
n	= numerosità dei valori argomentali

FREQUENZE

frequenza assoluta: $\sum_{i=1}^{n} fa_i = N$ frequenza relativa: $f_i = \frac{fa_i}{N}$

$$\sum_{i=1}^{n} f_i = \sum_{i=1}^{n} \frac{fa_i}{N} = \frac{1}{N} \sum_{i=1}^{n} fa_i = \frac{N}{N} = 1$$

$$fac_i = \sum_{k=1}^{i} fa_k \qquad fac_n = \sum_{k=1}^{n} fa_k = N$$

$$F_i = \sum_{k=1}^{i} f_k \qquad F_n = \sum_{k=1}^{n} f_k = 1$$

ISTOGRAMMA

$$h_i = \frac{f_i}{\Delta x_i} \qquad A = \sum_{i=1}^{n} h_i \cdot \Delta x_i = \sum_{i=1}^{n} \frac{f_i}{\Delta x_i} \cdot \Delta x_i = \sum_{i=1}^{n} f_i = 1$$

INDICI DI UNA DISTRIBUZIONE
DATI NON CUMULATI

$$a \le x_i \le b \quad \forall i$$

media $m_x = \frac{1}{N} \sum_{i=1}^{N} x_i = M[X]$ $\quad Na = \sum_{i=1}^{N} a \le \sum_{i=1}^{N} x_i \le \sum_{i=1}^{N} b = Nb$ (momento semplice di ordine 1)

$$a \le m_x \le b$$

con trasformazioni lineari: $y_i = a x_i + b$ $\qquad M[Y = aX + b] = aM[X] + b$

$$m_Y = \frac{1}{N_Y} \sum_{i=1}^{N_Y} y_i = \frac{1}{N_X} \sum_{i=1}^{N_X} (a x_i + b) = \frac{a}{N} \sum_{i=1}^{N} x_i + b = a m_X + b$$

mediana $med = \frac{x_{N/2} + x_{N/2+1}}{2}$ **(N pari)** $\qquad med = x_{(N+1)/2}$ **(N dispari)**

moda **= valore che si presenta il maggior numero di volte**

scarto $v_i = x_i - m_x$

media scarti $\frac{1}{N} \sum_{i=1}^{N} (x_i - m_x) = \frac{1}{N} \sum_{i=1}^{N} x_i - \frac{1}{N} \sum_{i=1}^{N} m_x = m_x - m_x = 0$

varianza (media dei quadrati degli scarti)

$s_x^2 = \frac{1}{N} \sum_{i=1}^{N} v_i^2 = M_x[V^2] \qquad S_x^2 \ge 0$ (momento centrale di ordine 2)

deviazione standard (scarto quadratico medio) $s_x=\sqrt{s_x^2}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-m_x)^2}$

varianza di un campione come stima della varianza dell'intera popolazione $\bar{s}_x^2=\frac{N}{N-1}s_x^2$

varianza $\boldsymbol{s_x^2=\frac{1}{N}\sum_{i=1}^{N}(x_i-m_x)^2}=\frac{1}{N}\sum_{i=1}^{N}(x_i^2-2x_i m_x+m_x^2)=$

$$=\frac{1}{N}\sum_{i=1}^{N}x_i^2-2m_x\frac{1}{N}\sum_{i=1}^{N}x_i+m_x^2\frac{1}{N}\sum_{i=1}^{N}1=\frac{1}{N}\sum_{i=1}^{N}x_i^2-2m_x^2+m_x^2\frac{1}{N}N=\boldsymbol{\frac{1}{N}\sum_{i=1}^{N}x_i^2-m_x^2}$$

Proprietà $s_X^2=M_X[V^2]$

se $y_i=x_i+b$ $\quad s_Y^2=\frac{1}{N_Y}\sum_{i=1}^{N_Y}(y_i-m_y^2)=\frac{1}{N_X}\sum_{i=1}^{N_X}(x_i+\not{b}-m_X-\not{b})^2=\frac{1}{N}\sum_{i=1}^{N}(x_i-m_X)^2=s_x^2$

se $y_i=a x_i+b$ $\quad s_Y^2=\frac{1}{N_Y}\sum_{i=1}^{N_Y}(y_i-m_y^2)=\frac{1}{N_X}\sum_{i=1}^{N_X}(a x_i+\not{b}-a m_X-\not{b})^2=\frac{1}{N}\sum_{i=1}^{N}a^2(x_i-m_X)^2=a^2 s_x^2$

Indice di asimmetria o **skewness** $\beta=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i-m_x}{s_x}\right)^3$

se $y_i=a x_i+b$ $\quad \frac{y_i-m_Y}{s_Y}=\frac{a x_i+b-a m_X-b}{|a|s_X}=\frac{\pm x_i-m_X}{s_X}$ $\quad \beta_Y=\pm\beta_Y$

DATI CUMULATI

media $m_x=\frac{1}{N}\sum_{i=1}^{N}x_i=\frac{1}{N}\sum_{j=1}^{n}\bar{x}_j fa_j=\sum_{j=1}^{n}\bar{x}_j\frac{fa_j}{N}=\sum_{j=1}^{n}\bar{x}_j f_j$ $\quad \sum_{i=1}^{n}\bullet f_i=M[\bullet]$

varianza $s_x^2=\frac{1}{N}\sum_{i=1}^{n}(\bar{x}_i-m_x)^2 fa_i=\sum_{i=1}^{n}(\bar{x}_i-m_x)^2 f_i=\sum_{i=1}^{n}\bar{x}_i^2 f_i-m_x^2$

scarto quadratico $s_x=\sqrt{\sum_{i=1}^{n}(\bar{x}_i-m_x)^2 f_i}=\sqrt{\sum_{i=1}^{n}\bar{x}_i^2 f_i-m_x^2}$

skewness $\beta=\frac{1}{N}\sum_{i=1}^{n}\left(\frac{\bar{x}_i-m_x}{s_x}\right)^3 f_i$

QUANTILI

$\xi_q \quad (0<q<1)$

se Nq **è intero** $\quad \xi_q=\frac{X_{Nq}+X_{Nq+1}}{2}$

se Nq **non è intero** $\quad \xi_q=X_{[Nq]+1}$ $\quad$ **con** $[Nq]$ **=parte intera**

Quantile 0,25	Percentile 25%	Primo quartile	Q_1	
Quantile 0,50	Percentile 50 %	Secondo quartile	Q_2	***MEDIANA***
Quantile 0,75	Percentile 75 %	Terzo quartile	Q_3	

MOMENTI

momento semplice di ordine n $\quad m_n=\frac{1}{N}\sum_i x_i^n$

momento centrale di ordine n $\quad \bar{m}_n=\frac{1}{N}\sum_i (x_i-m_x)^n$

VARIABILI DOPPIE

frequenza assoluta congiunta $\sum_{i=1}^{n}\sum_{j=1}^{m} fa_{ij}=N$

frequenza relativa congiunta $\sum_{i=1}^{n}\sum_{j=1}^{m} f_{ij}=1$

frequenza assoluta cumulata $fac_{kh}=\sum_{i=1}^{k}\sum_{j=1}^{h} fa_{ij}$

frequenza relativa cumulata $F_{kh}=\sum_{i=1}^{k}\sum_{j=1}^{h} f_{ij}$

istogramma $h=f_{ij}/(\Delta x_i \Delta y_i)$

$x \backslash y$	y_1	y_2		y_j		y_m	
x_1							
x_2							
x_i				i, j			q_i
x_n							
				r_j			

frequenza assoluta marginale $qa_i=\sum_{j=1}^{m} fa_{ij}$ $\sum_{i=1}^{n} qa_i=\sum_{i=1}^{n}\sum_{j=1}^{m} fa_{ij}=N$

frequenza relativa marginale $q_i=\sum_{j=1}^{m} f_{ij}$ $\sum_{i=1}^{n} q_i=\sum_{i=1}^{n}\sum_{j=1}^{m} f_{ij}=1$

frequenza assoluta marginale $ra_j=\sum_{i=1}^{n} fa_{ij}$ $\sum_{j=1}^{m} ra_j=\sum_{j=1}^{m}\sum_{i=1}^{n} fa_{ij}=N$

frequenza relativa marginale $r_j=\sum_{i=1}^{n} f_{ij}$ $\sum_{j=1}^{m} r_j=\sum_{j=1}^{m}\sum_{i=1}^{n} f_{ij}=1$

INDICI DI UNA DISTRIBUZIONE DOPPIA
Media

$$m_X=\sum_{i=1}^{n} x_i q_i \qquad m_Y=\sum_{j=1}^{m} y_j r_j$$

$$m_X=\sum_{i=1}^{n} x_i q_i=\sum_{i=1}^{n} x_i \sum_{j=1}^{m} f_{ij}=\sum_{i=1}^{n}\sum_{j=1}^{m} x_i f_{ij} \qquad m_Y=\sum_{j=1}^{m} y_j r_j=\sum_{j=1}^{m} y_j \sum_{i=1}^{n} f_{ij}=\sum_{j=1}^{m}\sum_{i=1}^{n} y_j f_{ij}$$

$$M_{XY}[\bullet]=\sum_{i=1}^{m}\sum_{j=1}^{n} \bullet\, f_{ij}$$

$$M_{XY}\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} M_{XY}[X] \\ M_{XY}[Y] \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n}\sum_{j=1}^{m} x_i f_{ij} \\ \sum_{i=1}^{n}\sum_{j=1}^{m} y_j f_{ij} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} x_i q_i \\ \sum_{j=1}^{m} y_j r_j \end{bmatrix} = \begin{bmatrix} M_X[X] \\ M_Y[Y] \end{bmatrix} = \begin{bmatrix} m_X \\ m_Y \end{bmatrix}$$

$$s_X^2 = \sum_{i=1}^{n} (x_i - m_X)^2 q_i \qquad s_Y^2 = \sum_{j=1}^{m} (y_j - m_Y)^2 r_j$$

$$s_X^2 = \sum_{i=1}^{n} (x_i - m_X)^2 q_i = \sum_{i=1}^{n} (x_i - m_X)^2 \sum_{j=1}^{m} f_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{m} (x_i - m_X)^2 f_{ij}$$

$$s_Y^2 = \sum_{j=1}^{m} (y_j - m_Y)^2 r_j = \sum_{j=1}^{m} (y_j - m_Y)^2 \sum_{i=1}^{n} f_{ij} = \sum_{j=1}^{m}\sum_{i=1}^{n} (y_j - m_Y)^2 f_{ij}$$

$$s_{XY} = \sum_{i=1}^{n}\sum_{j=1}^{m} (x_i - m_X)(y_j - m_Y) f_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j f_{ij} - m_X m_Y$$

$$\underline{X} = \begin{bmatrix} X \\ Y \end{bmatrix} \qquad C_{XY} = \begin{Bmatrix} s_X^2 & s_{XY} \\ s_{XY} & s_Y^2 \end{Bmatrix} = M_{XY}[\underline{VV}^+]$$

$$\underline{VV}^+ = \begin{pmatrix} V_X \\ V_Y \end{pmatrix} (V_X \quad V_Y) = \begin{pmatrix} X - m_X \\ Y - m_Y \end{pmatrix} (X - m_X \quad Y - m_y) = \begin{pmatrix} (X - m_X)^2 & (X - m_X)(Y - m_y) \\ (Y - m_y)(X - m_X) & (Y - m_Y)^2 \end{pmatrix}$$

Studio del legame tra le componenti di una variabile doppia

$$f_{ij} = q_i r_j \qquad \forall i, j$$
$$f[Y = g(x_i) | X = x_i] = 1$$
$$f[Y \neq g(x_i) | X = x_i] = 0$$

$$g(X) = aX + b$$

$$r_{XY} = \frac{s_{XY}}{s_X s_Y} = \frac{\sum_{ij} (x_i - m_X)(y_i - m_Y) f_{ij}}{\sqrt{\sum_{ij} (x_i - m_X)^2 f_{ij}} \sqrt{\sum_{ij} (y_i - m_Y)^2 f_{ij}}}$$

Indipendenza: $s_{XY} = \sum_{ij} (x_i - m_X)(y_j - m_Y) q_i r_j = \sum_{i} (x_i - m_X) q_i \sum_{j} (y_j - m_Y) r_j = 0$

Dipendenza: $s_{XY} = \sum_{ij} (x_i - m_X)(y_i - m_Y) f_{ij} = \sum_{i} (x_i - m_X) a (x_i - m_X) f_{ij} = a s_X^2$

$s_Y^2 = a^2 s_X^2 \qquad s_Y = |a| s_x$

$r_{XY} = \dfrac{s_{XY}}{s_X s_Y} = \dfrac{a s_X^2}{|a| s_X^2} = \dfrac{a}{|a|} = \pm 1$

In casi intermedi: $-1 \leq r_{XY} \leq +1$

Linearmente invariante: $\xi = a X + b \qquad \eta = \gamma Y + \delta \qquad r_{\xi\eta} = \pm r_{XY}$

RETTA DI REGRESSIONE

L'indice di correlazione lineare

$$r_{XY}=\frac{s_{XY}}{s_X s_Y}$$

$$y_i=a\,x_i+b \qquad \forall\, i$$

Se $Y=a\,X+b$, $m_Y=a\,m_X+b$ e $Y-m_Y=a(X-m_X)$

$$s_{XY}=M_X[(X-m_X)(Y-m_Y)]=M_X[(X-m_X)a(X-m_X)]=a\,M_X[(X-m_X)^2]=a\,s_X^2$$

$$s_Y^2=M_X[(Y-m_Y)^2]=M_X[a^2(X-m_X)^2]=a^2 s_X^2$$

$$s_Y=\sqrt{s_Y^2}=|a|s_X \quad \Rightarrow \quad r_{XY}=\frac{a\,s_X^2}{|a|s_X^2}=\pm 1$$

$$\xi=a_1 X+b_1 \qquad m_\xi=a_1 m_X+b_1 \qquad \xi-m_\xi=a_1(X-m_X) \qquad s_\xi=|a_1|s_X \qquad s_{\xi\eta}=a_1 a_2 s_{XY}$$

$$\eta=a_2 Y+b_2 \qquad m_\eta=a_2 m_Y+b_2 \qquad \eta-m_\eta=a_2(Y-m_Y) \qquad s_\eta=|a_2|s_Y \qquad r_{\xi\eta}=\pm r_{XY}$$

$$v_i=y_i-a\,x_i-b$$

$$\sum_{i=1}^{N} v_i^2=\sum_{i=1}^{N}(y_i-a\,x_i-b)^2=\min_{a,b}$$

$$y=\frac{s_{XY}}{s_X^2}x+m_Y-m_X\frac{s_{XY}}{s_X^2}$$

$$y-m_Y=\frac{s_{XY}}{s_X^2}(x-m_X)$$