INDICI DI VARIABILITA’

Definizione di VARIABILITA': la variabilità si può definire come l'attitudine di un carattere ad assumere diverse modalità quantitative. La variabilità è la quantità di dispersione presente nei dati.

Indici di variabilità assoluti o relativi

Indici di variabilità assoluti:

A) Intervalli di Variazione:
- **Range o Campo di variazione**
- **Differenza interquartile**

B) Scarti da un valore medio:
- **Scarto Quadratico Medio (e Varianza)**

C) Differenze Medie (cenni)

Indici di variabilità relativi:

- **Indici di variabilità relativi ad un valore medio;**
- Indici di variabilità relativi al massimo.

A) Intervalli di Variazione

Definizione di *RANGE* (ASSOLUTO) O CAMPO DI VARIAZIONE: Il *Range* è dato dalla differenza fra il valore massimo della distribuzione e il valore minimo.

$$\mathbf{R = x_{max} - x_{min}}$$

Range relativo alla media:

$$R_R = \frac{x_{\max} - x_{\min}}{m}$$

Range Medio:

Per dati semplici

$$R_m = \frac{x_{\max} - x_{\min}}{n-1}$$

Per dati ponderati

$$R_m = \frac{x_{\max} - x_{\min}}{N-1}$$

DIFFERENZA INTERQUARTILE

$$D = Q_3 - Q_1$$

ESERCIZIO

Nella tabella che segue vengono indicate le distribuzioni per classi di età dei suicidi e dei tentativi di suicidio verificatisi nel 1987 nell'ambito della popolazione femminile (fonte ISTAT 1988).
Confrontare i Quartili delle due distribuzioni.

età X	SUICIDI f(x)	TENTATIVI di SUICIDIO f'(x)	F(x)	F'(x)
0-13	3	5	3	5
13-17	19	101	22	106
17-24	46	236	68	342
24-44	259	526	327	868
44-64	434	321	761	1189
64-85	421	142	1182	1331
TOTALI	1182	1331		

Per la distribuzione dei SUICIDI:

Q_1=42; Q_2=56; Q_3=70 Q_4=85

D=70-42=28

Per la distribuzione dei TENTATIVI di SUICIDIO:

Q_1=24; Q_2=36; Q_3=52; Q_4=85

D=52-24=28

B) Scarti da un valore medio

$$_{A}S_{r} = \sqrt[r]{M\left|x_i - A\right|^{r}}$$

dove A è un valore medio e viene detta *origine degli scarti*.

IN PARTICOLARE

Lo SCARTO QUADRATICO MEDIO (: s.q.m. oppure σ) si definisce come la media quadratica degli scarti dalla media aritmetica:

$$\textbf{s.q.m.(x)} = \boldsymbol{\sigma_x} = {}_{m}S_{2} = \sqrt{M\left(x_i - m\right)^2}$$

Il valore dello scarto quadratico medio esprime l'oscillazione media dei valori della distribuzione attorno alla loro media aritmetica.

Il valore dello scarto quadratico medio è espresso nell'unità di misura originaria dei dati.

Formule:

-nel caso di dati semplici:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - m\right)^2}{n}}$$

-nel caso di dati ponderati:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left[(x_i - m)^2 f_i\right]}{\sum_{i=1}^{n} f_i}}$$

Definizione di VARIANZA: La varianza (: V(x) oppure σ^2) è la media del quadrato degli scarti di ciascuna osservazione dalla media aritmetica:

$$V(x)=\sigma^2=M[(x-m)^2]$$

Ovvero: la VARIANZA coincide con il radicando della formula dello Scarto Quadratico Medio.

Quindi la Varianza risulta espressa in termini quadratici.

<u>Metodo diretto</u> per il calcolo della V(X):

-nel caso di dati semplici:

$$V(X) = \sigma^2 = \frac{\sum_{i=1}^{n}(x_i - m)^2}{n}$$

-nel caso di dati ponderati:

$$V(X) = \sigma^2 = \frac{\sum_{i=1}^{n} \left[(x_i - m)^2 f_i \right]}{\sum_{i=1}^{n} f_i}$$

ESERCIZIO

Sui seguenti valori

x	f(x)
1	20
2	80
4	50
6	30
7	10

Determinare la Varianza di X.

SOLUZIONI

x	f(x)	x*f(x)	$(x-3,32)^2 * f(x)$
1	20	20	107,648
2	80	160	139,392
4	50	200	23,12
6	30	180	215,472
7	10	70	135,424
	190	630	621,056

M.Aritmetica=630/190=3,32

V(x)=621,056/190=3,269

ESERCIZIO

Sui seguenti valori

x	f(x)
x≤ 2	5
2<x≤ 4	60
4<x≤ 6	355
6<x≤ 8	70
x>8	10

a) Calcolare la media aritmetica della variabile X (per x≤2 porre x=1; per x>8 porre x=10);
b) Calcolare la Varianza di X.

SOLUZIONI

x	f(x)	x^c	xc*f(x)	$(x-5,1)^2*f(x)$
x≤2	5	1	5	84,05
2<x≤4	60	3	180	264,6
4<x≤6	355	5	1775	3,55
6<x≤8	70	7	490	252,7
x>8	10	10	100	240,1
	500		2550	845

a)
M.Aritmetica=2550/500=5,1

b) V(x)=845/500=1,69

<u>I Metodo indiretto</u> per il calcolo della V(X): la varianza si può ottenere più agevolmente come differenza fra la media dei quadrati di X e il quadrato della media aritmetica di X:

$$\mathbf{V(x)=\sigma^2=M(x^2)-m^2}$$

Formule:

-per dati semplici:

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

-per dati ponderati:

$$V(x)=\frac{\sum_{i=1}^{n}x_i^2 f_i}{\sum_{i=1}^{n}f_i}-\left(\frac{\sum_{i=1}^{n}x_i f_i}{\sum_{i=1}^{n}f_i}\right)^2$$

ESERCIZIO

Sui seguenti valori

x	f(x)
1	20
2	80
4	50
6	30
7	10

Determinare la Varianza di X.

SOLUZIONI

x	f(x)	x*f(x)	x^2*f(x)
1	20	20	20
2	80	160	320
4	50	200	800
6	30	180	1080
7	10	70	490
	190	630	2710

M.Aritmetica=630/190=3,32

Var(x)= $M(x^2)-(m_x)^2$= $2710/190-(3{,}32)^2$= 14,26-11,0224= 3,2376≈ 3,24

ESERCIZIO

Sui seguenti valori

x	f(x)
$x\leq 2$	5
$2<x\leq 4$	60
$4<x\leq 6$	355
$6<x\leq 8$	70
$x>8$	10

c) Calcolare la media aritmetica della variabile X (per $x\leq 2$ porre x=1; per x>8 porre x=10);
d) Calcolare la Varianza di X.

SOLUZIONI

x	f(x)	x^c	xc*f(x)	$(x^c)^2$*f(x)
x≤2	5	1	5	5
2<x≤4	60	3	180	540
4<x≤6	355	5	1775	8875
6<x≤8	70	7	490	3430
x>8	10	10	100	1000
			2550	13850

a)
M.Aritmetica=2550/500=5,1

b) $V(x)= M(x^2)-(m_x)^2= 13850/500-(5,1)^2= 1,69$

<u>II Metodo indiretto</u> per il calcolo della Varianza (*Proprietà pitagorica della varianza*):

$$V(x)=\sigma_x^2=\sigma_A^2-\varepsilon^2=M(x-A)^2-(A-m)^2$$

ESERCIZIO

Sui seguenti valori

x	f(x)
1	20
2	80
4	50
6	30
7	10

Determinare la Varianza di X con il II metodo indiretto, con A=3.

SOLUZIONI

x	f(x)	x*f(x)	$(x-3)^2*f(x)$
1	20	20	80
2	80	160	80
4	50	200	50
6	30	180	270
7	10	70	160
	190	630	640

$V(x)=640/190-(3-3{,}32)^2=3{,}36842-0{,}1024=3{,}266$

ESERCIZIO

Sui seguenti valori

x	f(x)
$x\leq 2$	5
$2<x\leq 4$	60
$4<x\leq 6$	355
$6<x\leq 8$	70
$x>8$	10

a) Calcolare la media aritmetica della variabile X (per $x\leq 2$ porre x=1; per x>8 porre x=10);

b) Calcolare la Varianza di X con il II metodo indiretto, con A=7.

SOLUZIONI

x	f(x)	x^c	xc*f(x)	$(x-7)^2*f(x)$
$x\leq 2$	5	1	5	180
$2<x\leq 4$	60	3	180	960
$4<x\leq 6$	355	5	1775	1420
$6<x\leq 8$	70	7	490	0
$x>8$	10	10	100	90
	500		2550	2650

a) $M(x)=2550/500=5{,}1$
b) $V(x)=2650/500-(7-5{,}1)^2=5{,}3-(1{,}9)^2=5{,}3-3{,}61=1{,}69$

LA STANDARDIZZAZIONE

Per standardizzazione si intende la procedura con cui una variabile *X* di media *m* e s.q.m. σ viene trasformata in una nuova variabile *U* di media pari a zero e s.q.m. unitario.
Analiticamente la standardizzazione consiste nella seguente trasformazione lineare:

$$u = \frac{x - m}{\sigma}$$

Indici di Variabilità RELATIVI

Indici di Variabilità <u>relativi ad un valore medio</u>

Definizione di COEFFICIENTE DI VARIAZIONE: Il coefficiente di variazione è dato dal rapporto fra lo scarto quadratico medio e la media aritmetica.

$$C_v = \frac{\sigma}{m}$$

Il coefficiente di variazione è una misura <u>relativa</u> di variabilità, <u>non</u> è espressa nell'unità di misura dei dati.

ESERCIZIO

In un comune la distribuzione di frequenze delle famiglie, classificate secondo il numero di figli X e il numero di stanze dell'appartamento abitato Y, è risultata la seguente:

	Y			
X	2	3	4	5
0	100	40	20	0
1	50	200	870	30
2	10	10	500	170

Confrontare la variabilità di X con quella di Y.

SOLUZIONI

L'Indice di Variabilità adatto a questo tipo di quesito è il *Coefficiente di Variazione* (***c.v.***): **c.v.=σ/m.**

Bisogna calcolarlo sia per X sia per Y:

X	f(x)	Y	f(y)
0	160	2	160
1	1150	3	250
2	690	4	1390
	2000	5	200
			2000

M(x)=1,265

Var(x)=1,955-$(1,265)^2$=0,354775

c.v.(x)=$\sqrt{0,354775}$ /1,265=**0,47085**

M(y)=3,815

Var(y)=0,5108

c.v.(y)=$\sqrt{0,5108}$ /3,815=**0,1874**

Dal confronto fra i valori dei due *c.v.* si deduce che la **X** è più variabile.