10/01/2024

martedì 9 gennaio 2024 12:40

§ 4. STATISTICA.

CHE COS'È LA STATISTICA? ELABORAZIONE E STUDIO DEI DATI RACCOLTI E COMPRENSIONE DI COME UN FENOMENO SI EVOLVERÀ (PREVISIONI).

CON L'INTRODUZIONE DEL CALCOLO DELLE PROBABILITÀ, SI È POTUTI PASSARE DALLA SOLA DESCRIZIONE DI DATI ALLA LORO GENERALIZZAZIONE.

LA STATISTICA CHE VEDREMO IN QUESTO CORSO SARÀ DI 2 TIPI:

1) DESCRITTIVA: STUDIO DI UN INSIEME DI DATI TRAMITE INDICI
2) INFERENZIALE: GENERALIZZAZIONE E APPROSSIMAZIONE PER FARE PREVISIONI.

OBIETTIVI DEL CORSO:

1) STUDIARE I PRINCIPALI INDICATORI USATI IN STATISTICA DESCRITTIVA

2) FARE MODELLI (LINEARI), QUANDO LE VARIABILI LO PERMETTONO, CHE LEGHINO 2 CARATTERI (STATISTICA BI-VARIATA)

ESEMPIO "TARGET": CONSIDERIAMO UN CAMPIO=
NE (= INSIEME DI DATI) DI 7 ALBERI DI CUI RIPORTIAMO ETÁ E DIAMETRO;

ETÁ (ANNI)	1	5	4	11	3	7	7
DIAM. (cm)	4	20	17	44	11	28	27

- CALCOLARE MEDIA E VARIANZA DELLA VARIABILE ETÁ
- CALCOLARE IL COEFFICIENTE DI CORRELAZIONE DEL DIAMETRO RISPETTO ALL'ETÁ E SCRIVER=
NE LA RETTA DI REGRESSIONE LINEARE.

STATISTICA DESCRITTIVA.

CI PONIAMO L'OBIETTIVO DI EVIDENZIARE LE CARATTERISTICHE DI UN INSIEME DI DATI SENZA GENERALIZZARNE LE PROPRIETÁ.

IL SINGOLO INDIVIDUO ALL'INTERNO DI UNA POPOLAZIONE É CHIAMATO UNITÁ STATISTICA.

IL SINGOLO INDIVIDUO ALL'INTERNO DI UNA POPOLAZIONE É CHIAMATO <u>UNITÁ STATISTICA</u>; LE PROPRIETÁ VERRANNO INDICATE CON LE LETTERE MAIUSCOLE $(X, Y, T, \dots)$, MENTRE CON LE MINUSCOLE I VALORI ASSUNTI IN CORRISPONDENZA DELL'UNITÁ STATISTICA.

AD ESEMPIO X POTREBBE ESSERE IL VOTO DEL DIPLOMA E $x_i = 99$ IL VOTO DELL'UNITÁ i-ESIMA SU UNA POPOLAZIONE DI $n = 25$ UNITÁ.

GLI x_i POSSONO ESSERE ORDINATI IN SENSO CRESCENTE CON LA NOTAZIONE

$$\underset{\text{VALORE MINIMO}}{x(1)}, \quad x(2), \quad \dots, \quad \underset{\text{VALORE MASSIMO}}{x(n)}$$

LE VARIABILI POSSONO ESSERE:

$$X = \begin{cases} \text{QUANTITATIVE (O NUMERICHE)} \\ \text{QUALITATIVE (O NON NUMERICHE)} \end{cases}$$

MENTRE LE VARIABILI QUANTITATIVE POSSONO ESSERE ORDINATE E HA SENSO PARLARE DI OPERAZIONI ARITMETICHE PER LE VARIABILI

ESSERE ORDINATE E HA SENSO PARLARE DI OPERAZIONI ARITMETICHE, PER LE VARIABILI QUALITATIVE CIÒ NON È POSSIBILE. AD ESEMPIO SE X = "COLORE DEGLI OCCHI" NON SI POSSONO ORDINARE I RISULTATI IN SENSO CRESCENTE.

IL NUMERO DI VALORI DIVERSI CHE PUÒ ASSUMERE UNA VARIABILE È DETTO <u>MODALITÀ</u> (AD ESEMPIO IL GRUPPO SANGUIGNO HA 4 MODALITÀ: 0, A, B, AB). LA PRIMA COSA CHE POSSIAMO CALCOLARE SU UNA POPOLAZIONE DI n UNITÀ CON VARIABILE X QUALITATIVA È LA <u>FREQUENZA</u> CON CUI CIASCUNA MODALITÀ SI VERIFICA: AVENDO $\{1, \dots, k\}$ MODALITÀ, CHIAMIAMO:

- FREQ. ASSOLUTA: N_J = n^o DI INDIVIDUI CON MODALITÀ J.
- FREQ. RELATIVA: $f_J = \frac{N_J}{n}$, DOVE $n = \#$ POPOLAZ.
- FREQ. PERCENTUALE: $p_J = (f_J \cdot 100)\ \%$

NOTA: IL CALCOLO DELLE FREQUENZE E LA DEFINIZIONE DI MODALITÀ VISTE SOPRA SI APPLICANO ANCHE A VARIABILI QUANTITATIVE. IN PARTICOLARE SE LE VARIABILI ...

ANCHE A VARIABILI QUANTITATIVE. IN PARTICOLARE, SE LE VARIABILI SONO CONTINUE POSSIAMO DIVIDERE IL CAMPO DEI VALORI IN INTERVALLI I_k E CONSIDERARE OGNI INTERVALLO COME UNA MODALITA'.

ESEMPIO:

	MASCHI	FEMMINE
CAMPIONE DI ZANZARE (SS)	12	188

LA VARIABILE X = GENERE HA 2 MODALITÁ, M ED F. LA MODALITA' M HA:

- FREQ. ASSOLUTA: 12
- FREQ. RELATIVA: $\frac{12}{200} = 0{,}06$
- FREQ. PERCENTUALE: $(0.06 \cdot 100)\% = 6\%$

ESERCIZIO: VIENE ESTRATTO IL DATO SUL COLORE DEGLI OCCHI DA UNA POPOLAZIONE DI $n = 500$ PERSONE:

	AZZURRO	CASTANO	VERDE	ALTRI
COLORE	89	342	65	4

CALCOLARE LA FREQ. ASSOLUTA, RELATIVA E PERCENTUALE PER CIASCUNA MODALITA'.

DEF: GLI <u>INDICI DESCRITTIVI</u> DANNO INDICAZIO= NI SULLA DISTRIBUZIONE DEI DATI. ESSI POSSONO ESSERE:

1. INDICI DI CENTRALITÀ (DANNO INFO. SULLA POSIZIONE CENTRALE ATTORNO ALLA QUALE SI DISTRIBUISCONO I DATI)
2. INDICI DI VARIABILITÀ (QUANTO I DATI SI ALLONTANANO DALLA POSIZ. CENTRALE)
3. INDICI DI FORMA / SIMMETRIA.

TRA GLI INDICI DI CENTRALITÀ ABBIAMO:

<u>MODA</u>: È LA MODALITÀ PIÙ FREQUENTE. ESSA VIENE CHIAMATA ANCHE "CLASSE MODALE". NON È ESCLUSO CHE CI SIANO PIÙ CLASSI MODALI.

<u>MEDIA ARITMETICA</u>: $\bar{x} = \frac{x_1 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$

CHIARAMENTE SE UN VALORE SI RIPETE, ESSO VA CONTATO TANTE VOLTE QUANTO LA SUA FREQUENZA

ES: $\{176, 181, 168, 176, 172\}$

$$\bar{x} = \frac{176 + 181 + 168 + 176 + 172}{5} = 174.6$$

$$\bar{x} = \frac{176+181+168+176+172}{5} = 174.6$$

MEDIANA: DATI n VALORI $x_1, \dots, x_n$, LI ORDINO IN SENSO CRESCENTE

$$x(1), \quad x(2), \quad \dots, \quad x(n)$$

SI CHIAMA MEDIANA IL VALORE $x_{0.5}$ (O $x_{50\%}$) CHE IN QUESTA SUCCESSIONE OCCUPA LA POSIZIONE CENTRALE, QUINDI SE n É DISPARI SI TRATTA DI $x\left(\frac{n+1}{2}\right)$, MENTRE SE n É PARI SI PRENDE LA MEDIA TRA $x\left(\frac{n}{2}\right)$, $x\left(\frac{n}{2}+1\right)$, QUINDI RICAPITOLANDO

$$x_{0.5} = \begin{cases} x\left(\frac{n+1}{2}\right) & \text{SE } n \text{ DISPARI} \\ \frac{1}{2}\left(x\left(\frac{n}{2}\right)+x\left(\frac{n}{2}+1\right)\right) & \text{SE } n \text{ PARI} \end{cases}$$

ESEMPI: (CALCOLARE MODA, MEDIA E MEDIANA)

1) $\{176, 181, 168, 176, 172\}$

MODA: 176

MEDIA: $\bar{x} = 174.6$

MEDIANA: PRIMA RIORDINIAMOLI...

$$\{168,\ 172,\ \boxed{176},\ 176,\ 181\}$$

$$\rightarrow x_{0.5} = 176$$

2)

ALTEZZA (cm)	166	168	169	170	171	172	173	174
FREQ. ASSOLUTA	1	3	6	11	8	6	4	3

MODA : ESERCIZIO

MEDIANA : ESERCIZIO

MEDIA :

$$\bar{x} = \frac{166 \cdot 1 + 168 \cdot 3 + 169 \cdot 6 + 170 \cdot 11 + 171 \cdot 8 + 172 \cdot 6 + 173 \cdot 4 + 174 \cdot 3}{1+3+6+11+8+6+4+3}$$

$$= \frac{166 + 404 + 1014 + 1870 + 1368 + 1032 + 692 + 522}{42}$$

CHECK!

$$= 170.7$$

3)

N° DI FIGLI PER FAMIGLIA	0	1	2	3	4	5	6
$F_{ASSOLUTA}$	6	12	16	9	4	1	2

ESERCIZIO : CALCOLARE IL N° MEDIO DI FIGLI PER FAMIGLIA.

IL PRINCIPALE INDICE DI VARIABILITA' É LA

VARIANZA: DATA X QUANTITATIVA, $x_1, \dots, x_n$ I VALORI ASSUNTI, $\bar{x}$ = MEDIA, ESSA È IL VALORE

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} = \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \dots + (x_n-\bar{x})^2}{n}$$

TALVOLTA DENOTATA CON σ_X^2 O CON $\mathrm{Var}(X)$

ESSA MISURA QUANTO I VALORI SI DISPERDONO (= SONO LONTANI) RISPETTO ALLA MEDIA.

DOMANDA: PERCHÉ INVECE DELLE DIFFERENZE AL QUADRATO $(x_i - \bar{x})^2$ NON ABBIAMO USATO LE DIFFERENZE "SEMPLICI" $(x_i - \bar{x})$?

$$\frac{\sum_{i=1}^{n} (x_i - \bar{x})}{n} = \frac{(x_1-\bar{x}) + (x_2-\bar{x}) + \dots + (x_n-\bar{x})}{n}$$

$$= \frac{x_1 + x_2 + \dots + x_n - n\bar{x}}{n} \qquad \bar{x} = \frac{\sum x_i}{n}$$

$$= \frac{\sum_i x_i - n \cdot \frac{\sum x_i}{n}}{n} = \frac{\sum x_i - \sum x_i}{n} = 0$$

GENERALIZZANDO LA FORMULA PER LA VARIANZA

SI DÁ LA SEGUENTE:

DEF: SI DEFINISCE <u>MOMENTO CENTRATO DI ORDINE k</u> IL VALORE

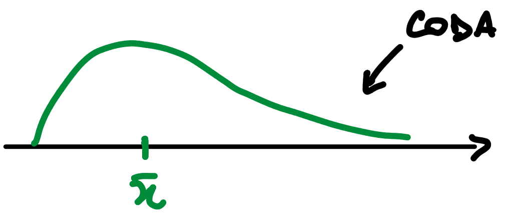
$$\mu_k = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^k}{n}$$

SI VEDE SUBITO CHE $\mu_2 = \sigma^2$

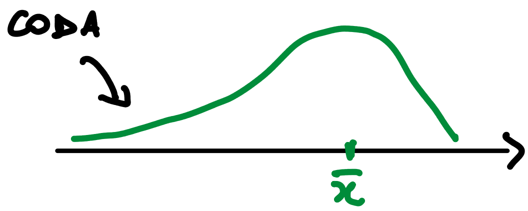
DA QUEST'ULTIMA OTTENIAMO ANCHE UN'INDICE DI FORMA/SIMMETRIA, L'<u>INDICE DI SKEWNESS</u>:

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} \qquad (\text{SE } \mu_2 = 0 \text{ SI PONE } \gamma_1 = 0)$$

SE $\gamma_1 > 0$ LA DISTRIBUZIONE DEI DATI É ASIMMETRICA CON CODA VERSO DX, PER $\gamma_1 < 0$ IL CONTRARIO



CASO $\gamma_1 > 0$



CASO $\gamma_1 < 0$

ESEMPIO: $\{168, 172, 176, 176, 181\}$

AVEVAMO $\bar{x} = 174.6$

s . ?

AVEVAMO $\bar{x} = 174.6$

$$\sigma^2 = \frac{\sum_{i=1}^{5}(x_i - \bar{x})^2}{5} =$$

$$= \frac{(168-174.6)^2+(172-174.6)^2+(176-174.6)^2+(176-174.6)^2+(181-174.6)^2}{5}$$

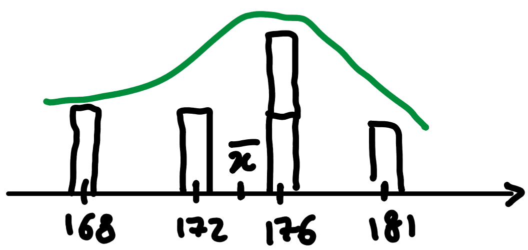
$$= \frac{(-6.6)^2+(-2.6)^2+(1.4)^2+(1.4)^2+(6.4)^2}{5} = \frac{95.2}{5} = 19.04$$

SKEWNESS : $\mu_3 = \frac{\sum_{i=1}^{5}(x_i - \bar{x})^3}{5} =$

$$= \frac{(-6.6)^3+(-2.6)^3+(1.4)^3+(1.4)^3+(6.4)^3}{5} = -7.488$$

$$\Rightarrow \gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{-7.488}{(19.04)^{3/2}} = \frac{-7.488}{83.08} = -0.09$$

$\gamma_1 < 0 \Rightarrow$ DISTRIBUZIONE ASIMMETRICA CON CODA VERSO SX .



ESERCIZIO: INVENTARE UN CAMPIONE CON $n \geq 5$, $\gamma_1 = 0$ E $\mu_2 \neq 0$.