DOMANDE D'ESAME (13 DOMANDE):

1)- CALCOLARE LA TRIMMED MEAN ALL'80%: 1, 1, 2, 3, 4, 5, 5, 6, 6, 50

2)- PER CONFRONTARE LA DISTRIBUZIOME DI UN CARATTERE SU DUE COLLETTIVI STATISTICI DI NUMEROSITÀ DIVERSA È OPPORTUNO CALCOLARE LE FREQUENZE ASSOLUTE, RELATIVE E PERCENTUALI. VERO O FALSO?

3)- INDICE V CRAMER (CALCOLARLO NELL'ESERCIZIO)

4)- LA TRIMMED MEAN ALL' 80% SIGNIFICA CALCOLARE LA MEDIA ESCLUDENDO IL 10% DEI VALORI PIÙ PICCOLI E IL 10% DEI VALORI PIÙ GRANDI. VERO O FALSO?

4)- NEL MODELLO DI REGRESSIONE $Y = a + bX + e$ UN VALORE POSITIVO MOLTO GRANDE DEL COEFFICIEMTE DI REGRESSIONE b INDICA:

A? C?

B? D?

5)- COS'È L'INDICE X^2? CALCOLARLO NEL SEGUENTE ESERCIZIO.

6)- COS'È IL COEFFICIENTE DI BRAVIS PEARSON?

7)- COME SI CALCOLA LA DIFFERENZA INTERQUARTILE?

8) FARE DELLE CONSIDERAZIONI (USANDO LE FREQUENZE %)

CLASSE ETÀ	POPOLAZIONE 1	POPOLAZIONE 2
DA 0-25	800	600
DA 25-50	1200	1200
DA 50-65	600	2000
DA 65 A 80	350	1000
MAGGIORE DI 80	50	200

9) MARIA PESA 85 Kg CON UNA DEVIAZIONE STANDARD PARI A 5 Kg E GIANNI PESA 60 Kg CON UNA DEVIAZIONE

STANDARD PARI A 7 Kg. QUAL È IL COEFFICIENTE DI VARIAZIOME PIÙ PICCOLO?

10)- IN UN GRAFICO P_P QUAMTO PIÙ I RESIDUI SI DISTRIBUISCONO NORMALMENTE TAMTO PIÙ I PUMTI SI TROVANO LONTANO DALLA BISETTRICE. VERO O FALSO?

11) UM COEFFICIEMTE DI DETERMINAZIONE R_2 PARI A 0,88 INDICA UM BUON ADATTAMENTO DELLA RETTA DI REGRESSIOME AI DATI CAMPIONARI.

12) PER TALI VALORI SI IMDICHI LA MEDIANA: 1, 1, 1, 2, 2, 2, 3, 3, 3, 5, 9, 10.

13) L'UNITÀ STATISTICA È UN'UNITÀ ELEMENTARE SU CUI VENGONO OSSERVATI I CARATTERI OGGETTO DI STUDIO.

14) LA SOMMA DELLE FREQUEMZE ASSOLUTE È SEMPRE PARI ALLA NUMEROSITÀ DEL COLLETTIVO. VERO O FALSO?

LA VARIABILITÀ: SINTESI DELLA DISTRIBUZIONE DI UN CARATTERE

A COSA SERVE? LA MEDIA È UN INDICE CHE FORNISCE UNA SINTESI DELLA DISTRIBUZIOME DEL CARATTERE OGGETTO DI STUDIO. PERÒ, LA DISTRIBUZIOME È BEN RAPPRESENTATA DALLA MEDIA QUANDO LA GRAN PARTE DELLE UNITÀ PRESENTANO UNA MODALITÀ VICINA ALLA MEDIA.

TRAMITE LA MEDIA ARITMETICA SI OTTIENE UNA INFORMAZIONE SINTETICA DELLA DISTRIBUZIONE PERDENDO PERÒ MOLTE INFORMAZIONI → TRA LE INFORMAZIONI PERDUTE C'È QUELLA SULLA VARIABILITÀ DEI DATI.

Ex: SI COMSIDERANO DUE GRUPPI DI STUDENTI E SI

MISURA LA LORO ALTEZZA.

CLASSE 1

STUDENTI	ALTEZZE
MARIO	1,77
GIOVANNI	1,78
LUCA	1,76
GAIA	1,76
MARIA	1,77
ILARIA	1,77
SERENA	1,78
FRANCESCO	1,77
ANTONIO	1,77
GIUSEPPE	1,77

MEDIA $\bar{x}$ = 1,77

CLASSE 2

STUDENTI	ALTEZZE
GIADA	1,61
MARINA	1,60
LUIGI	1,62
FILIPPO	1,60
ENRICO	1,60
ILARIA	1,89
SERENA	1,96
MARIO	1,97
FEDERICO	1,90
GIOVANNI	1,95

MEDIA $\bar{x}$ = 1,77

→ STESSA MEDIA ←

IN ENTRAMBI I CASI LA MEDIA ARITMETICA È UGUALE A 1,77. CM, MA LE POPOLAZIONI STATISTICHE NON SONO UGUALI.

↳ NEL PRIMO GRUPPO (CLASSE 1) GLI STUDENTI HANNO TUTTI UN'ALTEZZA VICINA ALLA MEDIA (1,77 CM).

↳ NEL SECONDO GRUPPO (CLASSE 2) GLI STUDENTI HANNO ALTEZZE MOLTO DIVERSE DALLA MEDIA (1,77).

LA MEDIA ARITMETICA È UN INDICE DI SINTESI CHE NON DÀ ALCUNA INFORMAZIONE SULLA VARIABILITÀ DEI DATI.

PER QUESTA RAGIONE È MOLTO UTILE ACCOMPAGNARE UN INDICE DI SINTESI CON UN INDICE DI VARIABILITÀ.

VARIABILITÀ: LA VARIABILITÀ DI UNA DISTRIBUZIONE ESPRIME LA TENDENZA DELLE UNITÀ DI UN COLLETTIVO AD ASSUMERE DIVERSE MODALITÀ DEL CARATTERE.

DUNQUE: LA DISPERSIONE O VARIABILITÀ DI UNA DISTRIBUZIONE DI DATI DEFINISCE LA FORMA PIÙ O MENO RACCOLTA DELLA DISTRIBUZIONE INTORNO AL VALORE CENTRALE.

COME SI MISURA LA VARIABILITÀ DI UNA DISTRIBUZIONE?

SI POSSONO UTILIZZARE DEGLI INDICI CHE SINTETIZZINO LA DIVERSITÀ TRA LE MODALITÀ ASSUNTE DAL CARATTERE E UN VALORE MEDIO (OPPURE LA DIVERSITÀ TRA DUE PARTICOLARI VALORI CARATTERISTICI DELLA DISTRIBUZIONE)

UN INDICE DI VARIABILITÀ DOVREBBE SODDISFARE ALMENO DUE REQUISITI:

- UN INDICE DI VARIABITÀ DEVE ASSUMERE IL SUO VALORE MINIMO SE E SOLO SE TUTTE LE UNITÀ DELLA DISTRIBUZIONE PRESENTANO UGUALE MODALITÀ DEL CARATTERE.

- UN INDICE DI VARIABILITÀ DEVE AUMENTARE ALL'AUMENTARE DELLA "DIVERSITÀ" TRA LE MODALITÀ ASSUNTE DALLE VARIE UNITÀ.

SE SI CONSIDERANO CARATTERI QUANTITATIVI, LA "DIVERSITÀ" TRA UNA MODALITÀ E UN VALORE CARATTERISTICO DELLA DISTRIBUZIONE (O TRA DUE VALORI CARATTERISTICI) VIENE MISURATA CONSIDERANDO O IL VALORE ASSOLUTO O IL QUADRATO DELLA LORO DIFFERENZA.

VARIANZA = σ^2

↳ È UN IMDICE BASATO SULLO SCOSTAMENTO DALLA MEDIA ARITMETICA. LA VARIANZA È LA MEDIA DEI QUADRATI DEGLI SCARTI DALLA LORO MEDIA ARITMETICA.

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

IL NUMERATORE DELLA VARIANZA È DETTO DEVIANZA $\sum_{i=1}^{m} (x_i - \bar{x})^2$

DISTRIBUZIONE PER UNITÀ

EX: DISTRIBUZIONE DI 5 ALUNNI PER PESO

1	65
2	72
3	48
4	81
5	64
TOTALE	330

$$\bar{x} = \frac{330}{5} = 66$$

$$\sigma^2 = \frac{(65-66)^2 + (72-66)^2 + (48-66)^2 + (81-66)^2 + (64-66)^2}{5}$$

$$\sigma^2 = \frac{1 + 36 + 324 + 225 + 4}{5} = \frac{590}{5} = 118$$

DISTRIBUZIOMI DI FREQUENZE

$$\sigma^2 = \frac{1}{m} \sum_{j=1}^{K} (x_j - \bar{x})^2 m_j$$

VOTO (x_i)	STUDENTI (n_i)
18	4
20	5
21	3
22	4
24	4
25	3
26	2
27	3
28	2
30	1
TOTALE:	31

MEDIA ARITMETICA PONDERATA:

$$\frac{(18\times4)+(20\times5)+(21\times3)+(22\times4)+(24\times4)+(25\times3)+(26\times2)+(27\times3)+(28\times2)+(30\times1)}{31}$$

$$\bar{x}=\frac{72+100+63+88+96+75+52+81+56+30}{31}=\frac{713}{31}=23$$

$$\sigma^2=\frac{(18-23)^2\times4+(20-23)^2\times5+(21-23)^2\times3+(22-23)^2\times4+(24-23)^2\times4+(25-23)^2\times3+(26-23)^2\times2+(27-23)^2\times3+(28-23)^2\times2+(30-23)^2\times1}{31}$$

$$\sigma^2=11,03$$

VARIANZA NELLE DISTRIBUZIONI PER CLASSI

CLASSI DI X	VALORI CENTRALI M_i	FREQUENZE
0 - 2	1 $(0+2)/2$	25
3 - 4	3,5 $(3+4)/2$	20
5 - 7	6 $(5+7)/2$	40
8 - 10	9 $(8+10)/2$	15
TOTALE		100

$$\bar{x} = \frac{(25 \times 1) + (20 \times 3,5) + (40 \times 6) + (15 \times 9)}{25 + 20 + 40 + 15} = \frac{25 + 70 + 240 + 135}{100} = \frac{470}{100} = 4,7$$

$$\sigma^2 = \frac{(1-4,7)^2 \times 25 + (3,5-4,7)^2 \times 20 + (6-4,7)^2 \times 40 + (9-4,7)^2 \times 15}{100}$$

$$\sigma^2 = 7,16$$

DEVIANZA

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

↳ NUMERATORE DELLA VARIANZA.

LA DEVIANZA O SOMMA DEI QUADRATI (SQ) DEGLI SCARTI DALLA MEDIA (SS = SUM OF SQUARES) È LA BASE DELLE MISURE DI DISPERSIONE DEI DATI, UTILIZZATE IN TUTTA LA STATISTICA PARAMETRICA.

TUTTA LA STATISTICA PARAMETRICA È FONDATA SULLA DEVIANZA E SULLE MISURE DA ESSA DERIVATE.

IL VALORE DELLA DEVIANZA DIPENDE DA DUE CARATTERISTICHE DELLA DISTRIBUZIONE:

- GLI SCARTI DI OGNI VALORE DALLA MEDIA (MISURA DELLA DISPERSIONE O VARIABILITÀ DEI DATI, EFFETTO CHE SI

INTENDE STIMARE).

- IL NUMERO DI DATI → FATTORE LIMITANTE PER L'USO DELLA DEVIANZA. UN CONFRONTO TRA DUE O PIÙ DEVIANZE RICHIEDEREBBE CAMPIONI CON LO STESSO NUMERO DI DATI

PER UNA MISURA DI DISPERSIONE DEI DATI CHE SIA INDIPENDENTE DAL NUMERO DI OSSERVAZIONI, SI RICORRE ALLA VARIANZA.

(SD O ROOT MEAN SQUARE DEVIATION) DEVIAZIONE STANDARD σ (O SCARTO QUADRATICO MEDIO)

↳ È DATA DALLA RADICE QUADRATA DELLA VARIANZA.

$$\sigma = \sqrt{\sigma^2}$$

↳ È UNA MISURA DI DISTANZA DALLA MEDIA E QUINDI HA SEMPRE UN VALORE POSITIVO.

↳ È UNA MISURA DI QUANTO I DATI SI DISTRIBUISCONO INTORNO ALLA MEDIA.

IL COEFFICIENTE DI VARIAZIONE CV

↳ DELLA DISTRIBUZIONE DI UN CARATTERE X, DI MEDIA $\bar{x} > 0$ E DEVIAZIONE STANDARD σ, È DATO DAL RAPPORTO TRA LA DEVIAZIONE STANDARD E LA MEDIA MOLTIPLICATO PER 100.

$$CV = \frac{\sigma}{\bar{x}} \cdot 100$$

OSSER... INDICI... DI GRA... E PESO... PERTAN... DIVERS...

· IL... · IL... DEL... · PU... È P... · N... BI...

OSSERVAZIONE: LA VARIANZA E LA DEVIAZIONE STANDARD SONO INDICI CHE RISENTONO DELL'UNITÀ DI MISURA E DELL'ORDINE DI GRANDEZZA DEI DATI (EX: PESO DELLE MADRI ADULTE E PESO DEI NEONATI).
PERTANTO IL CONFRONTO DELLA VARIABILITÀ TRA COLLETTIVI DIVERSI O VARIABILI DIVERSE RISULTA COMPROMESSO.

↓

IL COEFFICIENTE DI VARIAZIONE CV È UTILE PER CONFRONTARE LA VARIABILITÀ DI DUE DISTRIBUZIONI.

- IL SUO MINIMO È UGUALE A 0
- IL SUO MASSIMO NON È DEFINITO, VARIA AL VARIARE DEL TIPO DI DISTRIBUZIONE.
- PUÒ ESSERE UTILIZZATO PER STABILIRE SE UNA DISTRIBUZIONE È PIÙ VARIABILE DELL'ALTRA.
- NON DÀ INFORMAZIONI CIRCA L'INTENSITÀ DELLA VARIABILITÀ (OVVERO SE SIAMO VICINI AL SUO MASSIMO).

CAMPO DI VARIAZIONE O RANGE

↳ DATO UN INSIEME DI m VALORI OSSERVATI $X_1, X_2, \dots X_m$, ORDINATI IN SENSO CRESCENTE, DEFINIAMO CAMPO DI VARIAZIONE LA DIFFERENZA TRA IL PIÙ GRANDE E IL PIÙ PICCOLO DI TALI VALORI.

$$R = X_m - X_1$$

- NON È UN VERO E PROPRIO INDICE DI VARIABILITÀ (NON NECESSARIAMENTE CRESCE AL CRESCERE DELLA DIVERSITÀ DEI TERMINI)
- DI FACILE CALCOLO MA POCO USATO (DIPENDE SOLO DAI VALORI ESTREMI DELLA DISTRIBUZIONE ED È SUFFICIENTE LA PRESENZA DI UN SOLO DATO ERRATO PER MODIFICARNE