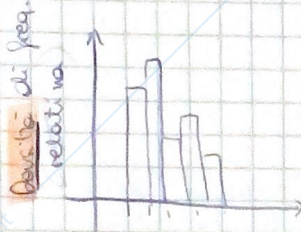


LEZIONE 6

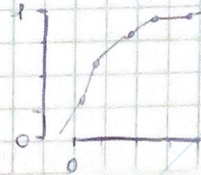
Rapp. grafiche (parte 2)

Rappresentazioni grafiche di distribuzioni per variabili in classi

Istogramma



Funzione di ripartizione empirica



Salta tramite segmenti di retta (incremento del fenomeno)
(usavano molto questo) $\text{MAX} = 1$

• INTRODUZIONE INDICE di POSIZIONE

Primo inx di indici che servono a sintetizzare l'info di una variabile x

In particolare indica la posizione della variabile all'interno dell'inx di riferimento

SCOPO • sintetizzare (in un unico valore l'informazione che proviene dalle variabili)

• localizzarla nell'inx dei valori di riferimento

Un valore che "sintetizza" l'informazione che subito un'idea dell'ordine di grandezza di un fenomeno

distingueremo 2 tipi di medie

Media larga

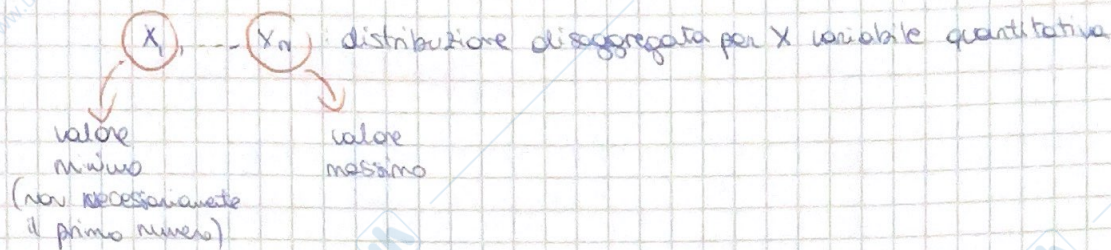
una delle modalità privilegiata o perché frequente, o perché in una particolare posizione (la sintesi meglio)

Media analitica

risultato di una manipolazione di dati

può non coincidere con le modalità osservate

NB: Ci sono dei minimi requisiti che una media deve soddisfare e essere tale



M è una generica media

deve soddisfare 3 requisiti:

① **Principio di Cauchy**: $X_1 \leq M \leq X_N$ deve essere compreso o eguale tra il valore + piccolo osservato e quello + grande

M è un valore che deve sintetizzare l'informazione della variabile

② **Principio di Chisini (o di invariante)**: $f(M, M, \dots, M) = f(X_1, X_2, \dots, X_N)$

La funzione applicata sui dati deve essere uguale (coincide) con la funzione quando a ciascun dato osservato sostituiamo la media

③ **Criterio delle minime perdite**: g (funzione generica) sintetizza la perdita di informazione che abbiamo passando da tutti i dati ad una singola media

M minimizza la perdita g

$$\sum_{i=1}^N (x_i - M)^2 \quad \text{se sono nel piano cartesiano}$$

MEDIE LASCHE

1° gruppo di indici di posizione

Tre tipi

- Moda
- Mediana
- Quartili e quantili

MODA

- Modalità che si presenta + spesso ← caratteristica + imp.
- Calcolabile x tutte le variabili
- può non essere unica → poco informativa

NB: la moda corrisponde alla modalità che si presenta con maggior frequenza

- Quando ci sono più mode si ha una distribuzione bimodale (meno informativa)

NB: Nella distribuzione in classi la moda corrisponde alla modalità che si presenta con più densità: $\left(\frac{n_i}{N}\right) \frac{1}{\Delta x_i}$ e si prende il valore di riferimento della classe

MEDIANA

• è il valore m associato all'unità statistica che occupa la posizione centrale della distribuzione
 → [prima bisogna ordinare le unità statistiche in ordine crescente rispetto alle modalità]

• le variabili devono essere **ordinabili** (le variabili qualitative su scala sconnessa non sono considerate)

• principio di Cauchy ($x_1 \leq m \leq x_n$)

• la somma dei valori assoluti degli scarti da un valore c è minima se $c = \text{mediana}$

$$\sum_{i=1}^N |x_i - m| = \min_c \sum_{i=1}^N |x_i - c|$$

⊛ Per distribuzioni di frequenza: $\frac{\text{tot. osservazioni}}{N}$ **DISPARI**

① ordinare le modalità in senso crescente

② si cerca l'unità che sta in posizione $(N+1)/2$

↓
 lascia alla sua sx e dx lo stesso numero di unità: $(N-1)/2$

③ la mediana è la **modalità** posseduta da quella unità ↑

⊛ Per distribuzioni di frequenza: N **PARI**

① ordinare le modalità in senso crescente

si cercano le unità in posizione $N/2$ e $N/2+1$ che lasciano la prima alla sua sx e la seconda alla sua dx lo stesso numero di ~~mediane~~ unità: $N/2-1$

↓
 caso di variabile qualitativa ordinale

↓
 - se la **modalità** posseduta dalle unità statistiche è la stessa, quella è la **MEDIANA**

- se la **più** modalità posseduta dalle unità Non è la stessa, la mediana non esiste

↓
 caso di variabile quantitativa discreta

↓
 - se la **modalità** posseduta dalle unità è la stessa, quella è la **MEDIANA**

- se la **modalità** posseduta dalle unità non è la stessa, ogni valore compreso tra le due modalità è **MEDIANA**, di solito si sceglie il valore centrale

⊛ Per distribuzioni in classi: individuazione della classe mediana $[c_{i-1}, c_i)$

poiché:

individuazione della mediana all'interno della classe mediana $[c_{i-1}, c_i)$

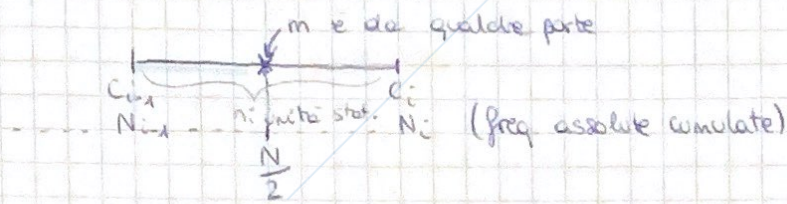
$$(m - c_{i-1}) : (N/2 - N_{i-1}) = (c_i - c_{i-1}) : n_i$$

$$\text{Si ha: } m = c_{i-1} + \frac{0.5 - F_{i-1}}{f_i} \text{ di}$$



⊛
 prendo dove F_i 50%!
 sopra

Intervallo generico: $[c_{i-1}, c_i)$



relazione: l'ampiezza di questo intervallo $[c_{i-1}, \frac{N}{2})$ ha un'ampiezza di questo tipo: $(m - c_{i-1})$ sta a tutte le unità statistiche che ci sono in quell'intervallo, come l'ampiezza di tutto l'intervallo sta a tutte le U.S. n_i .

$$(m - c_{i-1}) : (\frac{N}{2} - N_{i-1}) = c_i - c_{i-1} : n_i$$

unica quantità non nota

$$m - c_{i-1} = \frac{(\frac{N}{2} - N_{i-1})(c_i - c_{i-1})}{n_i}$$

0.5 = $\frac{1}{2}$ = $\frac{N}{2}$ freq. relativa cumulata = F_{i-1} ampiezza della classe = d_i

$$m = \frac{(\frac{N}{2} - N_{i-1})(c_i - c_{i-1})}{n_i} + c_{i-1}$$

$$m = c_{i-1} + \frac{(0.5 - F_{i-1}) \cdot d_i}{f_i}$$

(N3e) la mediana è un valore interno alla classe mediana

controllare sempre che il p.to sia all'interno della classe mediana

QUANTILE

- (estensione della mediana)
- indice associato a particolari posizioni nella distribuzione ordinate di X
- si parla di quantile di ordine q ($0 < q < 1$)
 ↓
 valore del supporto di X che suddivide la distribuzione in 2 parti
 ↓
 sinistra $q\%$ delle unità stat. destra $(1-q)\%$ delle unità stat.

⇒ • il quantile prende un nome particolare a seconda del valore di q del suo ordine che si considera

↓
QUARTILI (+ famosi) quantili con valori di $q = 1/4, 1/2, 3/4$
 (detti, ante primo, secondo, terzo quantile)

↓
 = MEDIANA (50%)

- si parla anche di **PERCENTILI** (q multiplo di 100)
- **DECILI** (q multiplo di 10)

13 • Dobbiamo avere delle **VARIABILI** almeno **ORDINABILI** x poter individuare i quantili

* **Quantili x variabili quantitative discrete**

- X con valori crescenti: $x_1, \dots, x_n$

- $h = \frac{2}{4} \cdot N$, $h = 1, 2, 3$

↓
 1, 2, 3
 quantili che
 identifica l'ordine del quantile ($1/4, 1/2, 3/4$)
 tot osservazioni

$$q_c = \begin{cases} \text{se la somma di } x \text{ a posizione } h \text{ e } x \text{ in posizione } h+1 & \text{se } h \text{ intero} \\ \frac{x_h + x_{h+1}}{2} & \text{per } h \text{ intero} \\ \text{oppure} \\ x_{[h]+1} & \text{per } h \text{ non intero} \end{cases}$$

$[h]$ = parte intera di h

Generale:

- Per trovare un **DECILE**: si suddivide la distribuzione in 10 parti e si cerca h

$$h = \frac{p}{10} \cdot N \quad p = 1, \dots, 9 \quad \text{poi si procede come per i quantili}$$

- Per trovare un **CENTILE**: si suddivide la distribuzione in 100 parti e si cerca h

$$h = \frac{p}{100} \cdot N \quad p = 1, \dots, 99 \quad \text{poi si procede come per i quantili}$$

* Quantili per distribuzioni in classi

- logica = per trovare la MEDIANA

→ questo valore cumulato è i quantili

$$m = C_{i-1} + \left(\frac{0,5 - F_{i-1}}{f_i} \right) \cdot d_i$$

per trovare la mediana

ES 1 cerchiamo l'ottavo decile (sx 80% dx 20%)
= quantile di ordine 0,8

① Si cerca sulla colonna delle freq. relative cumulate (F_i) dove superiamo l'80% delle U.S., prendiamo la propria CLASSE

② Si trova l'ottavo decile (d_8)
estremo inf. valore del quantile che stiamo cercando
 $d_8 = C_{i-1} + \left(\frac{0,8 - F_{i-1} (F_i \text{ della classe prec.})}{f_i \text{ (della classe del nostro decile)}} \right) \cdot d_i$

ampiezza classe

$$d_n = C_{i-1} + \left(\frac{v_n - F_{i-1}}{f_i} \right) \cdot d_i$$

NB 2 Il risultato deve rientrare nella classe da cui siamo partiti

ES 2 cerchiamo il primo quantile

$$q_n = C_{i-1} + \left(\frac{v_n - F_{i-1}}{f_i} \right) \cdot d_i$$