Ripasso sulle sommatorie

$\{x_i\}, i = 1, 2, \ldots, n$ Successione finita reale

$\sum_{i=1}^{n} \{x_i\}_{i=1,2,\ldots n} = x_1 + x_2 + x_3 \ldots x_n$ Somma delle componenti della successione

$\Downarrow$

In breve $\sum_{i=1}^{n} x_i$ sommatoria

Nota bene: *n* è la numerosità totale delle osservazioni

$$\sum_{i=1}^{n} x_i = \sum_{j=1}^{n} x_j \qquad x_1 + x_2 + x_3 + \dots x_n = x_1 + x_2 + x_3 + \dots x_n$$

$$\sum_{i=1}^{n} c = nc = c + c + c + c....$$ Costante *c* sommata n volte

$$\sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i$$

Teorema di linearità delle sommatorie: *a* non dipende dall'indice *i* e quindi lo posso "portare fuori" dall'operatore sommatoria

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

Media aritmetica

$$(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \ldots \tilde{x}_k)(f_1, f_2, f_3 \ldots f_k)$$

Successione di generiche modalità distinguibili (esempio delle età nella classe)

Distribuzioni di frequenza (frequenze assolute)

Importante: le modalità distinguibili sono *k* mentre la numerosità della popolazione è *n*. Ovviamente, $k<n$

$$\sum_{i=1}^{k} f_i = n$$

La somma delle frequenze assolute è la numerosità della popolazione

* **Nota bene: le frequenze assolute sono state indicate a lezione anche con n_i Qui si usa f_i coerentemente con la notazione che verrà usata più avanti durante il corso. La numerosità della popolazione è stata indicata con *n* ma verrà indicata anche con *N*.**

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$ **Media aritmetica (osservazioni)**

$$\mu = \frac{1}{n}\sum_{i=1}^{k} \tilde{x}_i f_i$$ **Media aritmetica (distribuzioni di frequenza)**

1) Proprietà di equilibrio $$\sum_{i=1}^{n}\left(x_i - \mu\right) = 0$$

2) Proprietà dei minimi quadrati $$\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \mu\right)^2 = \min \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \tau\right)^2$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \mu\right)^2 = \sigma^2$$ oppure $$\frac{1}{n}\sum_{j=1}^{k}\left(x_j - \mu\right)^2 f_j = \sigma^2$$ **Varianza**

Calcolo della varianza

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

$$\frac{1}{n}\sum_{i=1}^{n} x^2 + \frac{n}{n}\mu^2 - 2\mu\frac{\sum_{i=1}^{n} x_i}{n} = \frac{1}{n}\sum_{i=1}^{n} x^2 - \mu^2 = \sigma^2 = M(x^2) - \left[M(x)\right]^2$$

$$M(x^2) \geq \left[M(x)\right]^2$$

Definizione di varianza sfrutta la diseguaglianza di Jensen

- **Popolazione**: insieme di unità statistiche
- **Unità statistica**: ente elementare, supporto di una molteplicità di caratteri
- **Carattere**: classe di attributi associabile alle unità statistiche⇒caratteri quantitativi (variabili), caratteri qualitativi (mutabili)
- **Modalità**: elemento di una classe di attributi. Un carattere è definito dalla classe delle sue modalità

Esempio e scale

• Popolazione attiva di un comune	*popolazione*
• Attivo: nome e cognome	*unità statistica*
• Età	*carattere quantitativo*
• Anni:22	*modalità osservata*
• Sesso	*carattere qualitativo*
• Maschio	*modalità osservata*

- Caratteri qualitativi: scala sconnessa, scala ordinale
- Caratteri quantitativi: scala intervallare, scala rapporto

Indici di localizzazione

Esprimono l'ordine di grandezza del carattere

- **Moda**: è la modalità di un carattere che presenta frequenza massima (perde di significato quando il massimo non è unico)

- **Mediana**: modalità che lascia alla sua sinistra e alla sua destra $\frac{N-1}{2}$ osservazioni nel caso dispari e $\frac{N}{2}, \frac{N}{2}+1$ nel caso pari.

Importante: per il calcolo della mediana il carattere deve essere almeno ordinabile!!

Indici di mutabilità

- Distribuzione degenere: quando tutte le osservazioni assumono la stessa distribuzione il rischio previsivo è assente
- Mutabilità: attitudine del carattere ad assumere modalità distinte
- Misurazione della mutabilità: indici con alcune proprietà
- *assumono valore zero quando* $f_i = N, f_j = 0, \forall j \neq i$
- *assumono valore massimo in corrispondenza dell'equidistribuzione*
- *in situazioni intermedie assumono valori compresi tra gli estremanti esclusi* $0 < Z < max$

Indice di Gini

$$G = \sum_{i=1}^{k} p_i (1 - p_i) = 1 - \frac{1}{N^2} \sum_{i=1}^{k} f^2{}_i$$

$$0 \le G \le 1 - \frac{1}{K}$$

$$\tilde{G} = \frac{G - G_{\min}}{G_{\max} - G_{\min}} = \frac{G}{G_{\max}} = \sum_{j=1}^{k} p_j (1 - p_j) \frac{K}{K-1}$$

Indice di Shannon (Entropia)

$$H = -\sum_{i=1}^{K} p_i \ln p_i = \ln(N) - \frac{1}{N}\sum_{i=1}^{K} f_i \ln f_i$$

$$0 \leq H \leq \ln K$$

$$\tilde{H} = \frac{-\sum_{i=1}^{k} p_i \ln p_i}{\ln K}$$