Università di Pisa

Dipartimento di Informatica
Corso di Laurea Triennale in Informatica

Corso 2° anno - 6 CFU

Statistica

Professore:
Prof. Francesco Grotto

Autore:
Filippo Ghirardini

Anno Accademico 2023/2024

Contents

Statistica

Realizzato da: Filippo Ghirardini

A.A. 2023-2024

1 Statistica descrittiva

La statistica si occupa dello studio dei dati, ovvero della sua **raccolta**, **analisi** ed **interpretazione**. Le risposte dipendono dai dati e dalla conoscenza pregressa del problema, quindi da eventuali ipotesi ed assunzioni.

- Statistica **descrittiva**: quando i dati vengono analizzati senza fare assunzioni esterne per evidenziarne la struttura e rappresentarli in modo efficace
- **Inferenza statistica**: studia i dati utilizzando un modello probabilistico, ovvero supponendo che i dati siano valori assunti da *variabili aleatorie* con una certa *distribuzione di probabilità* dipendente da parametri non noti che devono essere stimati. Il modello potrà poi fare previsioni.

1.0.1 Campioni statistici

Definizione 1.0.1 (Popolazione)**.** *Insieme di oggetti o fenomeni che si vuole studiare su ognuno dei quali si può effettuare una stessa misura, ovvero un* ***carattere****. Può essere* ***ideale*** *o* ***reale****.*

Definizione 1.0.2 (Campione statistico)**.** *Un sottoinsieme della popolazione scelto per rappresentarla.*

Definizione 1.0.3 (Dati)**.** *Misure effettuate sul campione statistico.*

Definizione 1.0.4 (Frequenza)**.** *Può essere:*

- ***Assoluta****: il numero di volte in cui questo esito compare nei dati*
- ***Relativa****: frazione di volte in cui questo esito compare sul totale dei dati*

In generale dipendono dai dati e quindi non coincidono su tutta la popolazione.

Note 1.0.1. La scelta del campione in modo che sia rappresentativo è importante ma non verrà trattata.

1.0.2 Istogramma

Consiste in una serie di colonne ognuna delle quali ha per base un intervallo numerico e per area la frequenza relativa dei dati contenuti nell'intervallo.

Osservazione 1.0.1. La scelta delle ampiezze degli intervalli di base è cruciale. Un buon compromesso deve essere individuato sulla base della numerosità dei dati e sulla loro distribuzione.

Può avere varie forme:

- **Normale** se ha la forma di una *campana simmetrica*
- **Unimodale** se si concentra su una colonna più alta o **bimodale** se su due. Può essere asimmetrica a *destra* o a *sinistra* in base alla concentrazione dei dati in base al picco
- **Platicurtica** se i dati sono concentrati in un certo intervallo o **leptocurtica** se sono composti da un gruppo centrale e da molti *outliers*

1.0.3 Indici statistici

Dato un vettore $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ di dati numerici gli indici statistici sono quantità che riassumono alcune proprietà significative.

Definizione 1.0.5 (Media campionaria)**.** *La media aritmetica dei dati:*

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{1}$$

Definizione 1.0.6 (Mediana)**.** *Il dato x_i tale che la metà degli altri valori è minore o uguale ad esso e l'altra metà maggiore o uguale.*

Osservazione 1.0.2. La **mediana** è utile nel caso di dati molto **asimmetrici** ed è robusta rispetto alle code delle distribuzione. Al contrario la **media campionaria** viene facilmente spostata da dati molto piccoli o grandi.

Definizione 1.0.7 (Varianza campionaria)**.** *Si usa per misurare la dispersione dei dati attorno alla media campionaria.*

$$var(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{2}$$

È nulla se i dati sono tutti uguali. Possiamo mappare x diversamente:

- $x \mapsto x^2$ *misura la media dei punti della media campionaria*
- $x \mapsto x^3$ *misura la* ***sample skewness****, ovvero l'asimmetria della distribuzione*

$$b = \frac{1}{\sigma}\cdot\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3 \tag{3}$$

- $x \mapsto x^4$ *misura la piattezza della distribuzione dei dati, ovvero la* ***curtosi***

Definizione 1.0.8 (Scarto quadratico medio o deviazione standard)**.**

$$\sigma(x) = \sqrt{var(x)} \tag{4}$$

Proposizione 1.0.1. *Dato un campione di dati x ed un numero positivo d:*

$$\frac{\#\{x_i : |x_i - \bar{x}| > d\}}{n-1} \leq \frac{var(x)}{d^2} \tag{5}$$

Il termine a sinistra è la frazione di dati che differiscono dalla media campionaria più di d.

1.0.4 Quantili

Definizione 1.0.9 (Funzione di ripartizione empirica)**.** *Dato $x = (x_1, \dots, x_n) \in \mathbb{R}^n$:*

$$F_e(t) = \frac{\#\{i | x_i \leq t\}}{n} \tag{6}$$

Per ogni $t \in \mathbb{R}$ restituisce la frequenza relativa dei dati minori o uguali a t. È sempre ***non decrescente*** *e $F_e(-\infty) = 0$, $F(+\infty) = 1$.*

Definizione 1.0.10 (β-quantile)**.** *Il dato x_i tale che:*

- *almeno βn dati siano $\leq x_i$*
- *almeno $(1-\beta)n$ dati siano $\geq x_i$*

Inoltre:

- *Se βn non è intero vale $x_{(\lceil \beta n \rceil)}$*
- *Se βn è intero è la media aritmetica tra $x_{(\beta n)}$ e $x_{(\beta n+1)}$*

1.0.5 Dati multi-variati

Consideriamo coppie di dati **bivariati** del tipo

$$(x, y) = ((x_1, y_1), \dots, (x_n, y_n))$$

Definizione 1.0.11 (Covarianza campionaria)**.**

$$cov(x, y) = \sum_{i=1}^{n}\frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1} \tag{7}$$

Definizione 1.0.12 (Coefficiente di correlazione)**.** *Dati $\sigma(x) \neq 0$ e $\sigma(y) \neq 0$:*

$$r(x,y) = \frac{cov(x,y)}{\sigma(x)\sigma(y)} \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{t})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{8}$$

Misura la presenza di una relazione lineare tra i dati x e y quantificata dalla ***retta di regressione****.*

Proposizione 1.0.2 (Disuguaglianza di Cauchy-Scwarz)**.**

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \leq \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{9}$$

e quindi

$$|r(x,y)| \leq 1 \tag{10}$$

La **retta di regressione** è un'approssimazione dei dati con y_i con una combinazione lineare affine a $a + bx_i$, ottenuta cercando il minimo della distanza dai dati da questa retta con i quadrati degli scarti. L'obiettivo è quindi di cercare i parametri a e b calcolando

$$\inf_{a,b \in \mathbb{R}} \sum_{i=1}^{n}(y_i - a - bx_i)^2 \tag{11}$$

Teorema 1.0.1 (Retta di regressione)**.** Se $\sigma(x) \neq 0$ e $\sigma(y) \neq 0$, esiste un unico minimo al variare di $a, b \in \mathbb{R}$ della quantità 11, dato da:

$$b^\star = \frac{(n-1)cov(x,y)}{n \cdot var(x)} \qquad a^\star = -b^\star\bar{x} + \bar{y} \tag{12}$$

e vale

$$\min_{a,b \in \mathbb{R}} \sum_{i=1}^{n}(y_i - a - bx_i)^2 = (1 - r(x,y)^2)\sum_{i=1}^{n}(y_i - \bar{y})^2 \tag{13}$$

Quanto più $r(x,y)$ è vicino a 1, tanto più i valori tendono ad allinearsi con la retta. Se vale 1 vuol dire che i punti sono tutti sulla retta. Il segno di $r(x,y)$ corrisponde al segno del coefficiente angolare. Se è prossimo a zero allora non è una buona approssimazione.

2 Probabilità e indipendenza

La probabilità serve per quantificare l'incertezza misurando la fiducia che un evento possa accadere.

2.1 Spazi di probabilità

Definizione 2.1.1 (Spazio campionario)**.** *Lo spazio di probabilità* Ω *è l'insieme di tutti gli esiti possibili (**eventi elementari**)* ω *dell'esperimento. Ogni affermazione sulle misure corrisponde ad un sottoinsieme* $A \subset \Omega$ *degli esiti che la soddisfa. Ognuna delle affermazioni può essere combinata logicamente con le operazioni insiemistiche.*

Definizione 2.1.2 (Eventi incompatibili)**.**

$$A \cap B = \emptyset \tag{14}$$

Definizione 2.1.3 (Esperimento composto)**.** *Se un esperimento è composto da una successione ordinata di* n *sotto-esperimenti, il suo spazio campionario è*

$$\Omega = \{(\omega_1, \omega_2, \dots, \omega_n) | \omega_1 \in \Omega_1, \dots, \omega_n \in \Omega_n\} \tag{15}$$

dove Ω_i *è l'insieme degli esiti dell'i-esimo sotto-esperimento.*

Definizione 2.1.4 (σ-algebre)**.** *L'insieme di tutti i sottoinsiemi di* Ω *che sia chiuso per le operazioni logiche come **unione** e **intersezione**.*

Osservazione 2.1.1. Se due eventi sono incompatibili la probabilità che si realizzi uno qualsiasi dei due è la somma delle probabilità dei singoli eventi.

Definizione 2.1.5 (Probabilità)**.** *È il grado di fiducia che un evento si realizzi. È compreso tra* 0 *e* 1*. Più precisamente, dato* Ω *un insieme e* F *una* σ*-algebra di parti di* Ω*, è una funzione* $\mathbb{P} : F \to [0, 1]$ *tale che:*

- *l'evento certo ha probabilità* $\mathbb{P}(\Omega) = 1$
- *(**σ-addittività**) se* $(A_n)_{n=1,2,\dots}$ *è una successione di eventi a due a due **disgiunti**, vale*

$$\mathbb{P}\left(\bigcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} \mathbb{P}(A_n) \tag{16}$$

e nel caso di finiti sottoinsiemi disgiunti

$$\mathbb{P}\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{+N} \mathbb{P}(A_n) \tag{17}$$

Note 2.1.1. Si dice **trascurabile** un evento A tale che $\mathbb{P}(A) = 0$ e **quasi certo** un evento A tale che $\mathbb{P}(A) = 1$.

Proposizione 2.1.1. *Proprietà della probabilità:*

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ *e di conseguenza* $\mathbb{P}(\emptyset) = 0$
- $B \subset A \Longrightarrow \mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(B)$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- $\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C)$

Proposizione 2.1.2 (Limite di una successione di eventi)**.** *Data una successione di eventi* $A_1, \dots, A_n, \dots$*, questa può essere:*

- ***Crescente**:* $A_n \subseteq A_{n+1}$ *e quindi* $A = \bigcup_{n=1}^{+\infty} A_n = \lim_{n\to\infty A_n}$
- ***Decrescente**:* $A_n \supseteq A_{n+1}$ *e quindi* $A = \bigcap_{n=1}^{+\infty} A_n = \lim_{n\to\infty A_n}$

In entrambi i casi vale:

$$\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n) \tag{18}$$

2.2 Probabilità discreta

Definizione 2.2.1 (Probabilità discreta). *Dato* Ω ***numerabile***

$$\Omega = (\omega_1, \omega_2, \dots, \omega_n, \dots)$$

per ogni evento $A \subset \Omega$*, la misura di probabilità è:*

$$\mathbb{P}(A) = \sum_{\omega_i \in A} p_i = \sum_{\omega_i \in A} \mathbb{P}(\{\omega_i\}) \tag{19}$$

2.2.1 Probabilità uniforme su un insieme finito

Un esempio di probabilità discreta è quella uniforme su un insieme finito Ω, ovvero dove

$$p_1 = p_2 = \ldots = p_N$$

In questo caso vale:

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} = \frac{\text{"casi favorevoli"}}{\text{"casi possibili"}} \qquad A \subseteq \Omega \tag{20}$$

2.2.2 Calcolo combinatorio

Alcune formule notevoli:

- **Sequenze ordinate con ripetizione** di k numeri da 1 a n: n^k
- **Ordinamenti possibili** di $\{1, \dots, n\}$: $n!$
- **Sequenze ordinate senza ripetizione** di k numeri di $1, \dots, n$

$$\frac{n!}{(n-k)!} \qquad 0 \leq k \leq n$$

- **Sottoinsiemi** di $\{1, \dots, n\}$ formati da k elementi

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \qquad 0 \leq k \leq n$$

2.2.3 Funzione di massa

Definizione 2.2.2 (Funzione di massa). *Dato*

$$\Omega = \{x_1, x_2, \dots\} \subset \mathbb{R}$$

un sottoinsieme numerabile in cui ogni punto x_i *può contenere successioni (che possono andare a* $\pm\infty$*), la funzione di massa è*

$$\Omega \ni x_i \mapsto p(x_u) = \mathbb{P}(\{x_i\}) \in [0, 1] \tag{21}$$

Se poniamo che la probabilità di ogni altro punto non appartenente al sottoinsieme vale 0

$$x \neq x_i \implies p(x) = \mathbb{P}(\{x\}) = 0$$

allora possiamo estendere la funzione a $\mathbb{R}$ e dire che

$$\mathbb{P}(A) = \sum_{i : x_i \in A} p(x_i) \qquad \forall A \subseteq \mathbb{R} \tag{22}$$

Proposizione 2.2.1. *Valgono:*

$$p(x_i) \geq 0 \tag{23}$$

$$\sum_{i=1,2,\dots} p(x_i) = 1 \tag{24}$$

2.3 Probabilità condizionata

Quando si è a conoscenza della realizzazione di un evento, cambia la valutazione di probabilità di ogni altro evento.

Definizione 2.3.1 (Probabilità condizionata)**.** *Dati due eventi A, B con B non trascurabile, la probabilità condizionata di A rispetto a B è*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{25}$$

Proposizione 2.3.1 (Condizionamento ripetuto)**.** *Se l'intersezione di eventi $A_1 \cap \ldots \cap A_{n-1}$ non è trascurabile vale*

$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \ldots \cdot \mathbb{P}(A_n|A_1 \cap \ldots \cap A_{n-1}) \tag{26}$$

Definizione 2.3.2 (Partizione)**.** *Una partizione di Ω è una collezione di n eventi $B_1, \ldots, B_n$ a due a due disgiunti tali che*

$$B_1 \cup \ldots \cup B_n = \Omega \tag{27}$$

Definizione 2.3.3 (Sistema di alternative)**.** *È una partizione di Ω in eventi non trascurabili.*

Teorema 2.3.1 (Formula della probabilità o della fattorizzazione)**.** Dato $B_1, \ldots, B_n$ un sistema di alternative, per un qualunque evento A vale

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i) \tag{28}$$

Definizione 2.3.4 (Formula di Bayes)**.** *Dati A e B due eventi non trascurabili vale*

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)} \tag{29}$$

Definizione 2.3.5 (Formula di Bayes - Alternative)**.** *Dati A un evento e $B_1, \ldots, B_n$ un sistema di alternative vale*

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{j=1}^{n} \mathbb{P}(AB_j)\mathbb{P}(B_j)} \tag{30}$$

2.4 Indipendenza

L'idea è che la conoscenza che si è realizzato un certo evento non modifica la valutazione di probabilità di un altro evento.

Definizione 2.4.1. *Dati n eventi $A_1, \ldots, A_n$, questi sono indipendenti se per ogni k con $2 \leq k \leq n$ e per ogni scelta di interi $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ vale*

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdot \ldots \cdot \mathbb{P}(A_{i_k}) \tag{31}$$

Osservazione 2.4.1 (Complessità)**.** Il numero di uguaglianze da verificare per n eventi è

$$2^n - n - 1$$

Proposizione 2.4.1 (Spazi prodotto)**.** *Si consideri*

$$\Omega = \{a = (a_1, \ldots, a_n) | a_i = 0, 1\} = \{0, 1\}^n$$

su cui definiamo per ogni a la probabilità

$$\mathbb{P}(\{a\}) = p^{\#\{i:a_i=1\}}(a-p)^{\#\{i:a_i=0\}} = p^{\sum_{i=1}^n a_i}(a-p)^{n-\sum_{i=1}^n a_i}$$

E gli eventi

$$A_i = \{a \in \Omega : a_i = 1\} \qquad i = 1, \ldots, n$$

sono indipendenti tra di loro, così come i complementari A_i^c.

Osservazione 2.4.2. Due eventi possono essere indipendenti anche in presenza di una relazione causale. Viceversa due eventi possono essere dipendenti anche in assenza di una relazione causale.

2.5 Entropia di Shannon

Una misura di probabilità può essere uno strumento per quantificare l'informazione.

Definizione 2.5.1 (Entropia)**.** *Data una misura di probabilità discreta* $\mathbb{P}$ *su* $\Omega = \{x_1, \dots, x_n\}$*, con* $p_i = \mathbb{P}(\{x_i\})$*, la sua entropia è data dalla funzione*

$$H^{(n)}(p_1, \dots, p_n) = -\sum_{i=1}^{n} p_i \log(p_i) \tag{32}$$

Proposizione 2.5.1. *Valgono:*

1. *La funzione dell'entropia è **simmetrica**: scambiando* p_i *e* p_j *non cambia*
2. $H^{(n)}(1, 0, \dots, 0) = 0$
3. *È coerente tra* n *diversi:* $H^{(n)}(p_1 = 0, p_2, \dots, p_n) = H^{(n-1)}(p_2, \dots, p_n)$
4. $h^{(n)}(p_1, \dots, p_n) \leq H^{(n)}\left(\frac{1}{n}, \dots, \frac{1}{n}\right)$*, ovvero la massima entropia è data dalla distribuzione uniforme di probabilità*
5. *Data una probabilità su* $n \times m$ *oggetti* $\Omega = \{x_{11}, \dots, x_{ij}, \dots, xnm\}$ *con* $\mathbb{P}(x_ij) = q_{ij}$*, considerando gli eventi* $A_i = \{x_{i,1}, \dots, x_{i,m}\}$ *con* $\mathbb{P}(A_i) = p_1$ *vale*

$$H^{nm}(q_{11}, \dots, q_{ij}, \dots, q_{nm}) = H^{(n)}(p_1, \dots, p_n) + \sum_{i=1}^{n} p_i H^{(m)}\left(\frac{q_{i1}}{p_1}, \dots, \frac{q_{im}}{p_i}\right)$$

ovvero l'entropia è data da quella relative al sistema di alternative A_i *più la media pesata delle entropie relative nei blocchi* A_i*.*

Teorema 2.5.1 (Shannon)**.** Una funzione che soddisfa le 5 proprietà ha la forma

$$cH^{(n)} \qquad c > 0 \tag{33}$$

2.6 Densità di probabilità

Definizione 2.6.1 (Densità di probabilità)**.** *Una funzione non negativa* $f : \mathbb{R} \to [0, +\infty]$*, integrabile e tale che*

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

La sua probabilità è

$$\mathbb{P}(A) = \int_A f(x)dx \qquad A \subseteq \Omega \tag{34}$$

Osservazione 2.6.1. La probabilità di ogni singolo punto è nulla

$$\mathbb{P}(\{t\}) = \int_{\{t\}} f(x)dx = 0 \tag{35}$$

e in generale

$$\mathbb{P}(A) = 0 \qquad \forall A \subset \mathbb{R} \tag{36}$$

3 Variabili aleatorie

Le variabili aleatorie sono funzioni dello spazio di probabilità. Permettono di scrivere osservazioni diverse fatte su uno stesso spazio Ω.

Definizione 3.0.1 (Variabile aleatoria)**.** *È una funzione*

$$X : \Omega \to \mathbb{R} \tag{37}$$

definita su uno spazio di probabilità.

3.1 Legge du una variabile aleatoria

Ad una variabile aleatoria sono associati eventi del tipo "X prende valori in un insieme $A \subseteq \mathbb{R}$:

$$\{X \in A\} = X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$$

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A))$$