CONCETTI BASE DI STATISTICA

DEFINIZIONI

Probabilità

- Un numero reale compreso tra 0 e 1, associato a un evento casuale.
- Esso può essere correlato con la “frequenza relativa” o col “grado di credibilità” con cui un evento avviene.
- Per un alto grado di credibilità la probabilità è vicina al valore 1.
- Da un punto di vista non completamente corretto la probabilità può essere considerata come il rapporto tra il numero di eventi favorevoli e il numero degli eventi possibili nelle medesime condizioni.

Variabile aleatoria

- Una variabile aleatoria è una variabile che può assumere qualsiasi valore in un determinato intervallo, e alla quale è associata una distribuzione di probabilità (o densità di probabilità).
- Una variabile aleatoria che può assumere solo valori isolati è detta **variabile discreta**. Una variabile aleatoria che può assumere tutti i valori entro un intervallo finito o infinito è detta **variabile continua**.

Distribuzione di probabilità (di una variabile aleatoria)

- Una funzione che definisce la probabilità che una variabile aleatoria discreta assuma un determinato valore (o che una variabile aleatoria continua assuma tutti i valori di un intervallo).
- La probabilità che una variabile aleatoria possa assumere un qualsiasi valore tra quelli permessi è 1.

Densità di probabilità

- per una <u>variabile discreta</u>: una funzione che fornisce, per ogni valore x_i di una variabile aleatoria discreta X, la probabilità p_i che la variabile aleatoria si uguale a x_i. $p_i = Pr(X = x_i)$
- per una <u>variabile continua</u>: una funzione p(x) che fornisce, per ogni intervallo ($x \leftrightarrow x+dx$) dei valori che può assumere una variabile aleatoria continua X, la probabilità dP che la variabile aleatoria assuma un valore all'interno dell'intervallo. $dP = p(x)\,dx = Pr(x \leq X \leq x+dx)$
- La densità di probabilità coincide con la derivata (quando esiste) della funzione di distribuzione $p(x) = dP(x)/dx$

Normalizzazione della densità di probabilità

- Per il fatto che la probabilità che una variabile aleatoria possa assumere un qualsiasi valore tra quelli permessi vale 1, la densità di probabilità deve soddisfare a condizioni di normalizzazione:
- per una <u>variabile discreta</u>: Se i valori possibili sono (x_1, x_2, ….. ,x_N)

$$\sum_{i=1}^{N} p_i = 1 \qquad p_i \geq 0$$

- per una <u>variabile continua</u>: Se i valori possibili sono compresi nell'intervallo ($a \leftrightarrow b$)

$$\int_a^b p(x)\, dx = 1 \qquad p(x) \geq 0 \qquad a \leq x \leq b$$

Media o Valore Atteso

Per una variabile discreta: siano x_i i valori assunti dalla variabile aleatoria X con probabilità p_i . Il valore atteso, se esiste, risulta:

$$\mu = E(X) = \sum_i p_i \; x_i$$

la somma essendo estesa a tutti i valori x_i che può assumere la variabile X.

Per una variabile continua: sia p(x) la densità di probabilità associata alla variabile aleatoria X. Il valore atteso, se esiste, risulta:

$$\mu = E(X) = \int_D x \; p(x) \, dx$$

l'integrale essendo esteso a tutti gli intervalli che comprendono i possibili valori assunti da X.

Variabile aleatoria centrata

Una variabile aleatoria il cui valore atteso sia nullo. Se la variabile aleatoria X ha un valore atteso uguale a μ , la corrispondente variabile aleatoria centrata è (X - μ).

Varianza

La varianza di una variabile aleatoria, o di una distribuzione di probabilità, è il valore atteso del quadrato della corrispondente variabile centrata

$$\sigma^2 = V(X) = E\left\{\left[X - E(X)\right]^2\right\} = E\left(X^2\right) - \left[E(X)\right]^2$$

Deviazione standard

La deviazione standard di una variabile aleatoria, o di una distribuzione di probabilità, è la radice quadrata positiva della varianza

$$\sigma = \sqrt{V(X)} = \sqrt{E\left\{\left[X - E(X)\right]^2\right\}} = \sqrt{E\left(X^2\right) - \left[E(X)\right]^2}$$

Distribuzione Normale o Gaussiana

La distribuzione di probabilità di una variabile aleatoria X la cui densità di probabilità è

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty \le x \le \infty$$

μ è il valore atteso della variabile aleatoria X e σ è la relativa varianza.

mezzo di attributi).

Popolazione

La totalità degli elementi in considerazione. Nel caso di una variabile aleatoria, la distribuzione di probabilità (densità di probabilità) definisce la popolazione di quella variabile.

Frequenza

Il numero di volte in cui un dato tipo di evento si avvera. Dal punto di vista della misura : il numero di osservazioni che cadono in una specifica classe.

Distribuzione di frequenza

La relazione empirica tra i valori di una caratteristica e la loro frequenza (frequenza relativa). In questi casi la distribuzione può essere rappresentata in differenti modi, ad es. :

- per mezzo di un istogramma
- per mezzo di un grafico a barre
- per mezzo di una tabella a due entrate

Valor medio o valore atteso

Il valore atteso della variabile aleatoria z, rappresentato col simbolo μ_z e detto anche valor medio di z, è dato da

$$\mu_z \equiv E(z) = \int z \; p(z) \, dz$$

La sua stima statistica è data dalla media aritmetica dei valori z_i assunti dalla variabile z di densità di probabilità p(z).

$$\bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i$$

Varianza

La varianza di una variabile aleatoria coincide col valore atteso (medio) del quadrato della deviazione dal suo valore atteso. Quindi la varianza di una variabile aleatoria z, di densità di probabilità p(z), è data da:

$$\sigma^2(z) = \int (z - \mu_z)^2 \; p(z)\, dz$$

essendo μ_z il valore atteso di z. La varianza σ^2(z) può essere stimata da

$$s^2(z) = \frac{1}{n-1} \sum_{1=1}^{n} (z_i - \bar{z})^2 \qquad \bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i$$

Il fattore (n-1) nella espressione di s^2(z) proviene dalla correlazione tra i valori z_i e $\bar{z}$ e riflette il fatto che vi sono solo (n-1) termini indipendenti nel set di valori $\{z_i - \bar{z}\}$.

Se il valore atteso μ_z della variabile z è noto (non stimato) la varianza può essere stimata da

$$s^2(z_i) = \frac{1}{n} \sum_{i=1}^{n} (z_i - \mu_z)^2$$

La varianza della media aritmetica delle osservazioni, piuttosto che la varianza di una singola osservazione, è la misura appropriata dell'incertezza del risultato di una misurazione.
La varianza della variabile z, $\sigma^2(z)$, deve essere accuratamente distinta dalla varianza della media aritmetica .
La varianza della media aritmetica di una serie di n osservazioni indipendenti z_i della grandezza rappresentata dalla variabile aleatoria z è data da

$$\sigma^2(\bar{z}) = \sigma^2(z)/n$$

ed è stimata dalla varianza sperimentale della media:

$$s^2(\bar{z}) = \frac{s^2(z)}{n} = \frac{1}{n(n-1)} \sum_{i=1}^{n} (z_i - \bar{z})^2$$

Deviazione standard
La deviazione standard è la radice quadrata positiva della varianza.

GRADI DI LIBERTA' E LIVELLI DI CONFIDENZA

Il problema è di ottenere dalla stima y della grandezza misurata Y, e dalla sua incertezza $u_c(y)$, una **incertezza espansa** $U_p = k_p\, u_c(y)$ che definisca un intervallo $y\text{-}U \leq Y \leq y\text{+}U$ tale che abbia una elevata probabilità di copertura (o un elevato **livello di confidenza**) p.

Si deve quindi determinare il fattore di copertura k_p che genera un intervallo intorno al risultato y della misurazione che ci si aspetta contenga una grande, specifica frazione p della distribuzione di valori che potrebbero ragionevolmente essere attribuiti alla grandezza Y da misurare.

Per ottenere il fattore di copertura k_p che produce un intervallo corrispondente a uno specifico livello di confidenza p si richiede una dettagliata conoscenza della distribuzione di probabilità caratterizzata dai risultati della misura e dalla incertezza.

Ad esempio, per una grandezza z descritta dalla **distribuzione normale** di valor medio μ_z e deviazione standard σ_{z}, il valore di k_p che produce un intervallo $\mu_z \pm k_p \sigma_z$ che comprende una frazione p della distribuzione, può essere calcolato facilmente.

Alcuni esempi sono riportati in tabella.

Livello di confidenza p (in %)	*Fattore di copertura k_p*
68.27	1
90	1.645
95	1.960
95.45	2
99	2.576
99.73	3
100	∞