Ottavio Serra

Statistica descrittiva e Calcolo delle probabilità.

1. Elementi di statistica descrittiva.

La statistica ha due scopi principali:

I) Ricavare da un insieme di dati, troppo numerosi per essere esaminati singolarmente e proficuamente, alcune informazioni significative per il particolare problema da studiare.

II) Fornire metodi che servano ad *imparare dall'esperienza*, giustificando fin dove è possibile il passaggio da osservazioni particolari a leggi generali.

Nel primo caso si parla di statistica descrittiva, nel secondo di statistica induttiva o inferenziale. Quest'ultima ha bisogno di nozioni di calcolo delle probabilità, mentre la statistica descrittiva non ne fa uso esplicitamente. Partiamo perciò da questa, dando alcuni elementi di statistica inferenziale dopo aver introdotto concetti e tecniche di calcolo delle probabilità.

Nella statistica descrittiva si parte da una popolazione che può essere di vario tipo: le molecole di un gas, una coltura batterica, gli studenti di una scuola, i professori di una classe. Di questa popolazione si studiano alcuni caratteri o attributi e si ripartisce la popolazione in classi, a seconda dei caratteri. Un carattere può essere

(a) qualitativo, come per esempio il mezzo di locomozione usato dagli studenti per andare a scuola;

(b) ordinale come quando si fa una scala di preferenze o di simpatia;

(c) numerico. In tal caso l'attributo (o carattere) viene detto *variabile*.

Per un carattere di tipo (a) c'è poco da dire, l'unico dato significativo è quello di *classe modale*.

a) Moda o classe modale di una popolazione è la classe più numerosa. Per esempio, rispetto alla popolazione scolastica di una scuola, la Moda è la classe dei pedoni, se questi sono i più numerosi rispetto agli studenti che usano altri mezzi di locomozione.

b) Mediana. Se i valori della popolazione X sono ordinati, la *mediana* di X è quel valore (o quei valori) rispetto al quale X ammette tanti valori minori quanti sono i maggiori. Per esempio nella sequenza ordinata 44, 46, 50, 51, 59, la mediana è 50. (Potrebbe essere, in migliaia di euro, il reddito di 5 gruppi di persone: il reddito mediano è 50). La classe modale sarebbe il gruppo di persone col reddito più alto.

Supponiamo ora che in una popolazione S è assegnato un carattere espresso da una variabile numerica reale. In tal caso è determinata una funzione a valori reali X: S→***R.*** S potrebbe essere un gruppo di m operai classificati secondo le ore lavorate in un anno. Se x_i sono le ore dell'i^{mo} operaio (la variabile X assume i valori $x_1, x_2,.. x_m$) è solito chiamare la quantità $M(X) = \bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i$, *media aritmetica* di X.

c) Valore medio o **speranza matematica**: Se una variabile X può assumere i valori $x_1, x_2,..x_n$ con *pesi statistici* $p_1=m_1/m$, $p_2=m_2/m,...p_n=m_n/m$ $(p_1+p_2+p_n=1)$, la media pesata di X è M(X)= $\bar{x} = \sum_{i=1}^{n} p_i x_i$. Se $p_i=1/m$, $\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i$ (**media aritmetica**). Però rispetto alla media i valori di X possono essere più o meno concentrati o sparpagliati. Una misura della *concentrazione* può essere data dal valore medio dello scarto assoluto $s = \sum_{i=1}^{n} p_i \left|x_i - \bar{x}\right|$.

Non ha senso il valor medio dello scarto, perché questo è sempre zero;

infatti, $\sum_{i=1}^{n} p_i(x_i - \bar{x}) = \sum_{i=1}^{n} p_i x_i - \sum_{i=1}^{n} p_i \bar{x} = \bar{x} - \bar{x}\sum_{i=1}^{n} p_i = \bar{x} - \bar{x}.1 = 0$. Ma lo scarto assoluto è disagevole nei calcoli, perché il valore assoluto non è una funzione *liscia* (non è derivabile), perciò si usa lo scarto quadratico medio (o deviazione standard) σ definito come la radice quadrata della *varianza*: $Var(X) = \sum_{i=1}^{n} p_i (x_i - \bar{x})^2 = \sum_{i=1}^{n} p_i x_i^2 - 2\bar{x}\sum_{i=1}^{n} p_i x_i + \bar{x}^2 = M\left(x^2\right) - \bar{x}^2$, da cui $\sigma = \sqrt{Var(X)}$. Si noti che $M(x^2)$ è sempre maggiore o uguale di $\bar{x}^2$,perché…

Accanto alla varianza di una variabile, è usata anche la *Covarianza* di due variabili X e Y definita come valore medio del prodotto degli scarti: $Cov(X,Y) = \sum_{i=1}^{n} p_i(x_i - \bar{x})(y_i - \bar{y})$. Sviluppando il prodotto si ottiene:

$$Cov(X,Y) = \sum_{i=1}^{n} p_i x_i y_i - \bar{x}\sum_{i=1}^{n} p_i y_i - \bar{y}\sum_{i=1}^{n} p_i + \bar{x}.\bar{y}\sum_{i=1}^{n} p_i = M(XY) - \bar{x}.\bar{y}.$$

La retta di regressione o dei minimi quadrati. A volte si sospetta che ci sia un legame statistico tra due variabili di una stessa popolazione, che sia approssimativamente lineare. Per esempio, statura e peso per gli studenti di una scuola, della stessa età e sesso. Se si rappresentano i valori di X (statura) sull'asse delle ascisse e i valori di Y (peso) sull'asse delle ordinate, non si otterrà praticamente mai una retta, per le fluttuazioni statistiche. Si cerca allora la retta di equazione y = ax+b che meno si discosta dai dati misurati, cioè si determinano i parametri a e b in modo che sia minimo lo scarto quadratico medio dei punti (x_i; y_i) dalla retta y=ax+b (retta di regressione). Lo scarto quadratico medio è la radice quadrata della varianza rappresentata dalla funzione $F(a,b) = \frac{1}{m}\sum_{i=1}^{m}(y_i - ax_i - b)^2$. Per minimizzare F rispetto ai parametri, occorre annullare le derivate di F: F'_a ed F'_b :

$$\begin{cases} \sum_{i=1}^{m}(y_i - ax_i - b)x_i = 0 \\ \sum_{i=1}^{m}(y_i - ax_i - b) = 0 \end{cases} \Rightarrow \begin{cases} M(XY) - aM(x^2) - b\bar{x} = 0 \\ \bar{y} - a\bar{x} - b = 0 \end{cases} \Rightarrow \begin{cases} b = \bar{y} - a\bar{x} \\ M(XY) - aM(x^2) - (\bar{y} - a\bar{x})\bar{x} = 0 \end{cases}.$$

Dalla 2ª si ricava $a = \frac{M(XY) - \bar{x}.\bar{y}}{M(x^2) - \bar{x}^2} = \frac{Cov(X,Y)}{Var(X)} = \frac{Cov(X,Y)}{(\sigma_X)^2}$ e dalla 1ª si ricava b.

Si chiama coefficiente di correlazione tra X e Y il rapporto $r = \frac{Cov(X,Y)}{\sigma_X . \sigma_Y}$, che è una specie di coseno (varia tra -1 e 1); Le variabili sono tanto più correlate quanto più r è prossimo ad 1 (in valore assoluto); se r è prossimo a zero, X ed Y si dicono *non correlate*; se r=0, X e Y si dicono indipendenti o *ortogonali.*

2. Spazi di probabilità discreti. .

Sia S un insieme, finito o numerabile, detto *Spazio degli Eventi;* A, B sottoinsiemi, detti *Eventi*; introdotta una *misura* m su S, chiameremo probabilità di un evento A il rapporto tra m(A) e m(S). p(A)=m(A)/m(S).

La probabilità di un evento è un numero p compreso tra 0 e 1, che gode delle seguenti proprietà:

1° $p(A \cup B) = p(A)+p(B)- p(A\cap B)$; Spiegare a livello intuitivo.
Se A e B sono disgiunti, $A\cap B=\Phi$, allora $p(A \cup B) = p(A)+p(B)$.

2° $p(A/B) = \frac{p(A\cap B)}{p(B)}$ La probabilità di A condizionata al verificarsi di B è la probabilità della parte di A inclusa in B relativamente a B, come se lo spazio degli eventi si *contraesse* in B: $p(A/B)= \frac{m(A\cap B)}{m(B)} = \frac{m(A\cap B)/m(S)}{m(B)/m(S)} = \frac{p(A\cap B)}{p(B)}$.

Per simmetria $p(B/A) = \frac{p(A\cap B)}{p(A)}$ e quindi

$$p(A\cap B) = p(B).p(A/B) = p(A).p(B/A).$$

A si dice indipendente stocasticamente da B (indipendenza in senso probabilistico) se p(A/B)=p(A); in tal caso $p(A\cap B)$=p(A).p(B), da cui deriva che p(B/A)=p(B): Se A è indipendente da B, B è indipendente da A. Perciò si dice che A e B sono (stocasticamente) indipendenti.

3° Se A_1, A_2, … A_n, …è una successione di eventi incompatibili (cioè a due a due disgiunti), allora

$$p(\bigcup_{k\in N} A_k) = \sum_{k=1}^{\infty} p(A_k) \quad \text{(additività numerabile)}$$

La probabilità dell'unione numerabile di insiemi disgiunti (cioè di eventi incompatibili) è la somma della serie delle singole probabilità. E' chiaro che la serie deve convergere a un numero (non negativo e) non maggiore di 1.

4° Se due eventi A e B sono incompatibili, cioè se il verificarsi di uno esclude il verificarsi dell'altro, e inoltre sono complementari (uno dei due si deve verificare), allora p(A)+p(B)=1.

[p+q=1]

Esempio1. In un'urna ci siano 7 gettoni bianchi e 3 neri. Se riteniamo che ogni gettone ha la stessa probabilità di essere estratto, (distribuzione ***uniforme***) allora p=p(bianco)=7/10 e a q=p(nero)=3/10. p+q=1. Se però nell'urna ci sono anche 8 gettono rossi, p+q= 7/18+3/18 = 10/18 <1.

Esempio2. Un tiratore ha probabilità p=0,7 di colpir il bersaglio: che probabilità ha di colpire almeno una volta in 3 tiri?

I Metodo: P(almeno una volta su 3) = 1 – P(nessuna volta) = 1- $(0,3)^3$ = 0,973.

II Metodo: detta q la probabilità di non colpire (p+q=1 → q= 1-0,7 = 0.3), si ha

P(almeno una volta su 3) = (pqq+qpq+qqp)+(ppq+pqp+qpp)+(ppp) = $3(pq^2)+3(p^2q)+p^3$ =0,973

(**Meglio il primo metodo**!).

Esempio3. Lanciando due dadi i risultati vanno da 2 (=1+1) a 12 (=6+6). La somma 7 è la più probabile: come mai? Il primo a dare una giustificazione fu Galilei. (Si intende, dadi non truccati).

Esempio4. Due amici ugualmente bravi a briscola (o scopa) puntano somme uguali con l'accordo che vince l'intera posta chi per primo arriva a 5 vittorie. Le mogli però li interrompono quando il primo ha vinto 4 partite e il secondo 3. Come devono dividersi la posta? (Problema proposto dal cavaliere de Mèray a Pascal). Suggerimento: il primo vince il torneo se impedisce al secondo di vincere due partite consecutive: P(primo) = p+qp, P(secondo) = qq. Siccome p=q=1/2,
P(primo) =3/4 e P(secondo) = ¼, perciò la posta va divisa nella proporzione di 3 ad 1.
Generalizzare per p e q generiche e nell'ipotesi che al 1° manchino x partite e al 2° ne manchino y.

Esempio5. Qual'é la probabilità di fare ambo alla ruota di Napoli? I numeri sono 90 e se ne estraggono 5, perciò lo spazio degli eventi (lo spazio di tutte le cinquine) ha misura $C_{90,5} = \binom{90}{5} = \frac{90.89.88.87.86}{1.2.3.4.5}$. I casi favorevoli sono dati da quelle cinquine che contengono i numeri puntati, tolti i quali gli altri 3 possono essere qualunque, dunque la misura dell'evento è data dai casi favorevoli che sono $C_{88,3} = \binom{88}{3} = \frac{88.87.86}{1.2.3}$ e

P(ambo)= $\frac{88.87.86}{1.2.3}.\frac{1.2.3.4.5}{90.89.88.87.86} = \frac{4.5}{90.89} = \frac{2}{9.89} = \frac{2}{801}$.

Esercizio. Calcolare le probabilità P(terno), P(quaterna), P(cinquina).

Esempio6. Totocalcio. Supponendo di trascurare il fattore campo e la forza delle squadre, siccome ogni partita ha 3 risultati : 1, x, 2, la probabilità di azzeccare un risultato è p=1/3 e di sbagliarlo è perciò q= 2/3. La probabilità di imbroccare k risultati su 13 sarà P(k su 13)=$\binom{13}{k}(1/3)^k(2/3)^{13-k}$.

Verificare che $\sum_{k=0}^{13}\binom{13}{k}(1/3)^k(2/3)^{13-k}=1$.

Esercizio. Qual è la probabilità di fare almeno un punto sulla schedina? E di fare meno di 3 punti?

Esempio7. Lanciando una moneta 10 volte (è lo stesso che 10 monete simultaneamente), qual'é la probabilità di ottenere esattamente 3 Teste? E quella di ottenere almeno 3 Teste?

E' più probabile ottenere 10 Teste su 10 lanci, oppure 5 Teste seguite da 5 Croci?

Qual è la probabilità di ottenere 5 Teste e 5 Croci, se si prescinde dall'ordine?

Osservazione. In generale, se la probabilità di evento è p(E) = p e dell'evento contrario è q=1-p, la probabilità che E si verifichi k volte in una sequenza di n prove (**prove ripetute**)

è $P(n,k)=\binom{n}{k}p^kq^{n-k}=\frac{n!}{k!(n-k)!}p^kq^{n-k}$. **Distribuzione binomiale o bernoulliana.**

Se p=q=1/2, $P(n,k)=\frac{n!}{k!(n-k)!}p^kq^{n-k}=\frac{n!}{k!(n-k)!}\left(\frac{1}{2}\right)^n$. ($0 \le k \le n$).

In tal caso si dimostra che P(n,k) ha un massimo per k=n/2 se n è pari o due massimi consecutivi per k=(n-1)/2 e per k=(n+1)/2 se n è dispari.

In generale si ha il massimo (o i due massimi) per k compreso tra np-q ed np+p (estremi inclusi).

Se n è grande, la distribuzione di probabilità assume la forma a campana di Gauss e possiamo dire che la probabilità massima spetta al caso che l'evento E si presenti un numero di volte pari ad np (se np è intero, altrimenti all'intero più vicino) e di conseguenza che l'evento contrario si presenti nq volte. **In una sequenza di prove ripetute le probabilità di un evento e del suo contrario sono proporzionali alle rispettive probabilità.**

Esempio8. Lanciando una moneta 10 volte, l'evento più probabile è 5 Teste (e 5 Croci) **5=10.1/2.**

P(10;5)=$\frac{10!}{5!5!}\frac{1}{2^{10}}$=3628800/(120)^{2}1/1024=252/1024=**0,246.**

P(10;4)=P(10;6)= $\frac{10!}{4!6!}\frac{1}{2^{10}}$=3628800/(24x720)1/1024=210/1024=**0,205** < **0,246**.

Però la probabilità massima decresce al crescere d n; per esempio, lanciando la moneta 20 volte si ha **P(20;10)**=20!/(10!)^{2}1/2^{20}=184756/1048576=0,176 < **P(10;5)**; ma appena ci si allontana dal caso più probabile le probabilità sono ancora più piccole e decrescono rapidamente.

Però il calcolo esatto dei fattoriali, anche per numeri non troppo grandi, è molto laborioso. In tal caso si approssima n! con la formula di Stirling: **log(n!)=nlog(n)-n**, da cui $n! = Ae^{-n}n^n$, essendo A un fattore correttivo che vale circa $\sqrt{2\pi n}$.

Esempio9. Già conosciamo i concetti di valore medio e varianza. Basta riprendere quanto detto nella 1ª sezione (Statistica descrittiva) e interpretare i pesi statistici come probabilità. Lanciando un dado, il punteggio va da 1 a 6; si chiede il punteggio medio. La variabile aleatoria X assume i valori 1, 2,… 6 con probabilità costante p=1/6; $\bar{x}=\frac{1}{6}(1+2+3+4+5+6)=\frac{1}{6}\left(\frac{6*7}{2}\right)=\frac{1}{6}\left(\frac{42}{2}\right)=3,5$.

Il valor medio del quadrato è $M(X^2)=\frac{1}{6}(1^2+2^2+...+6^2)=\frac{91}{6}\approx 15,167$, perciò la varianza

$Var(X)=M(X^2)-\bar{x}^2=\frac{91}{6}-\left(\frac{7}{2}\right)^2=\frac{35}{12}\approx 2,917$

e lo scarto quadratico medio $\sigma_X=\sqrt{2,917}\approx 1,708$.

Esempio10. Si voglia calcolare il numero medio di tiri necessari per colpire un bersaglio, essendo p la probabilità costante di colpirlo in ciascun tiro.

Posto q=1-p, si ha: $\bar{x} = p + 2qp + 3q^2 p + 4q^3 p + ... + nq^{n-1} p + ... = p\sum_{n=1}^{\infty} nq^{n-1}$ Questa serie converge ad $1/(1-q)^2 = 1/p^2$ (**credetemi sulla parola**), perciò $M(X) \equiv \bar{x} = p\frac{1}{p^2} = \frac{1}{p}$. (Abbastanza intuitivo, a livello qualitativo). E la varianza? $Var(X) = M(X^2) - \bar{x}^2$ $=M(X^2)-1/p^2$, ma $M(X^2)$ è difficile da calcolare: $M(X^2) = p\sum_{n=1}^{\infty} n^2 q^{n-1} = ..??.. = \frac{2-p}{p^2}$. Segue $Var(X) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$.

(**Serie ancor più difficile della prima da calcolare. Sapreste scrivere un algoritmo per approssimarle?**).

Se, per esempio, p=0,2 si ottiene:

```
Media(X)   =p*Som(n=1..Inf)n*(1-p)^(n-1);
Media(X^2)=p*Som(n=1..Inf)n^2*(1-p)^(n-1);
Var(X)= Media(X^2)-(Media(X))^2.
 p (<0 o =0 per Fine)=0.2
Media(X)   =5.00000
Media(X^2)=45.000000
VAR(X)     =20.000001
______________________
```

Somma e prodotto di variabili aleatorie.

M(X+Y) =M(X)+M(Y). Dim:

M(X+Y) = $\sum_{i,k} p_{i,k}(x_i + y_k) = \sum_i p_i^X x_i + \sum_k p_k^Y y_k = M(X) + M(Y)$, cdv.

M(XY)=M(X).M(Y), se X e Y sono indipendenti. Dim: M(XY)=$\sum_{i,k} p_{i,k} x_i y_k = \sum_{i,k} p_i^X p_k^Y x_i y_k$ (per l'indipendenza stocastica) = $\sum_i p_i^X x_i \,.\, \sum_k p_i^Y y_k$ =M(X)M(Y).

Passiamo ora alla varianza. **Var(X+Y)=Var(X)+Var(Y)+2Cov(x,y).** Dim:

Var(X+Y)=$M((X+Y-M(X+Y))^2)$=$M((X+Y-\bar{x}-\bar{y})^2) = M((X-\bar{x})^2 + (Y-\bar{y})^2 + 2(X-\bar{x})(Y-\bar{y}))$; tenendo conto che il valor medio di una soma è la soma dei valori medi, segue la tesi. **La varianza della somma si riduce alla somma delle varianze se X e Y sono indipendenti**, cioè se Cov(X,y)=0.

Osserviamo ancora che M(λX)=λ.M(X) e che $Var(\lambda X)=\lambda^2 Var(X)$ da cui segue $\sigma(\lambda X)=|\lambda|\sigma(X)$.

UN CASO IMPORTANTE: valor medio e varianza della variabile **X = successo con probabilità p:**

P(x=1)=p, P(x=0)=q, [p+q=1]. $M(X) \equiv \bar{x} = p.1 + q.0 = p$.

$Var(X) = p(1-\bar{x})^2 + q(0-\bar{x})^2 = p(1-p)^2 + q(0-p)^2 = pq^2 + qp^2 = pq(p+q) = pq$. Lo scarto quadratico medio sarà $\sigma = \sqrt{pq}$.

Prove ripetute. Supponiamo ora di avere n variabili indipendenti X_1, X_2,… X_n con la stessa legge: stesso valore medio $\bar{x}$ e stessa varianza σ^2. Posto Y = $X_1+X_2+\ldots X_n$, avremo

$M(Y) = n\bar{x}$ e $Var(Y) = Var(X_1 + ... X_n) = n\sigma^2$, quindi $\sigma_Y = \sigma_X \sqrt{n}$. Se le n variabili contano i successi in n prove, M(Y) = np e $\sigma_Y = \sqrt{npq}$. Come si vede, la dispersione cresce indefinitamente con n. Però, se consideriamo la variabile Y/n, che conta la percentuale dei successi, cioè la frequenza relativa dei successi: ν/n, il suo valor medio è np/n = p e Var (Y/n)= σ^2/n =pq/n; perciò $\sigma_Y = \sqrt{\frac{pq}{n}}$, che diminuisce al divergere di n.

Teorema di Cebicev e legge dei grandi numeri di Jakob Bernoulli.

Sia Y la variabile aleatoria (ν/n –p) e y_1, y_2,…y_n i suoi valori; il valor medio dei quadrati è

$M(Y^2) = \sigma_r^2 = \sum_{i=1}^{n} p_i y_i^2$. Fissato un arbitrario numero reale positivo ε, alcuni degli y_i, diciamo i primi k, sono in modulo minori di ε, i rimanenti maggiori o uguali. Perciò $\varepsilon^2 \sum_{i=k+1}^{n} p_i \leq \sigma_r^2$. Segue che $\sum_{i=k+1}^{n} p_i \equiv P\left(\left|\frac{\nu}{n} - p\right| \geq \varepsilon\right) \leq \frac{\sigma_r^2}{\varepsilon^2}$ da cui $P\left(\left|\frac{\nu}{n} - p\right| < \varepsilon\right) > 1 - \frac{\sigma_r^2}{\varepsilon^2}$,cioè $P\left(\left|\frac{\nu}{n} - p\right| < \varepsilon\right) > 1 - \frac{pq}{n\varepsilon^2}$.

Questa disuguaglianza, trovata da Cebicev (1821-1894), giustifica la legge dei grandi numeri di Bernoulli (1654-1705):

La probabilità che in una serie di prove ripetute lo scarto tra la frequenza relativa e la probabilità a priori sia minore in valore assoluto di un arbitrario numero reale positivo ε tende ad 1 (alla certezza) al crescere del numero n delle prove. $\lim_{n \to \infty} P\left(\left|\frac{\nu}{n} - p\right| < \varepsilon\right) = 1$.

Ciò non significa che ν/n → p per n → ∞, ma solo che tende alla certezza la probabilità di uno scarto arbitrariamente piccolo.

Esempio 11. Quante volte devo lanciare una moneta perché la frequenza relativa dell'evento "Testa" sia compresa tra 0,49 e 0,51 con una probabilità maggiore del 90%? Si ha p=q=0,5, ε = 0,01; perciò $1 - \frac{1}{4n10^{-4}} = \frac{9}{10} \Rightarrow \frac{1}{4n10^{-4}} = \frac{1}{10} \Rightarrow n = 25000$. (Ci vogliono 25000 lanci).

Vedremo in un'altra lezione che si possono ottenere valutazioni più stringenti (valori di n più piccoli).