Prova scritta di "Analisi statistica dei dati sperimentali"
5/12/2002

Problema 1

La grandezza derivata $z = y \times \ln x$ viene calcolata in seguito alla misura delle grandezze x e y, i cui valori sono riportati in tabella:

$$x = 11, 13, 12, 13, 14, 13, 14$$
$$y = 37, 37, 36, 37, 35, 36, 34$$

a) Le grandezze x e y in base alle misure risultano correlate?
b) Quanto vale l'incertezza di $\bar{z}$?
c) Se z e' una temperatura, quali sono le dimensioni di x e y?
d) Dopo qualche tempo, viene effettuata una nuova misura e si ottengono i seguenti valori: $(x, y) = (17, 30)$. Possiamo sospettare della validita' (in parte o totale) di questa nuova misura? Commentare il risultato.

Soluzione

a) calcolo il valore del coefficiente di correlazione lineare r:

$$N = 7;\ \sum x_i = 90;\ \sum x_i^2 = 1164;\ \sum y_i = 252;\ \sum y_i^2 = 9080;\ \sum x_i y_i = 3235$$

$$r = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{[N \sum x_i^2 - (\sum x_i)^2]^{1/2} [N \sum y_i^2 - (\sum y_i)^2]^{1/2}} = -0.67$$

Considero ora l'ipotesi "i dati non sono correlati", e faccio il test sul valore del coefficiente di correlazione lineare:

$$P_N(|r| > 0.67) = 0.10 > 0.05$$

Il risultato implica che l'ipotesi non puo' essere rigettata (in altre parole i dati non sono correlati).

b) Posso pertanto procedere a trovare i momenti principali delle grandezze misurate x e y.

$$\overline{x} = 12.86;\ s_x = 1.07;\ s_{\overline{x}} = \frac{s_x}{\sqrt{N}} = 0.40$$

dove ho mantenuto temporaneamente un numero "alto" di cifre. Analogamente:

$$\overline{y} = 36.00;\ s_y = 1.16;\ s_{\overline{y}} = \frac{s_y}{\sqrt{N}} = 0.44$$

I valori finali sono:

$$\overline{x} = 12.9 \pm 0.4;\ \overline{y} = 36.0 \pm 0.4$$

$$\overline{z} = \overline{y} \ln \overline{x} = 92.1$$

$$s_{\overline{z}} = [(\partial z/\partial x)^2 s_{\overline{x}}^2 + (\partial z/\partial y)^2 s_{\overline{y}}^2]^{1/2} = 1.5$$

$$\overline{z} = 92.1 \pm 1.5$$

c) l'esistenza della funzione $\ln x$ implica che la grandezza x deve essere adimensionale in quanto essa e' sviluppabile in serie di Taylor contenente termini x elevati a potenze diverse, e non e' possibile sommare grandezze dimensionalmente diverse. Quindi $[z] = [y] = $ T [K].

d) Poiche' abbiamo ipotizzato x e y non correlate (non abbiamo cioe' potuto escludere l'ipotesi chye non lo fossero), l'aggiunta di una nuova misura ci impone, a priori, di trattarla esattamente come le altre misure fatte sinora. Debbo quindi inizialmente ricalcolare $\overline{x}, s_x, \overline{y}, s_y$ e successivamente applicare il criterio di Chauvenet alla misura "sospetta" (sia per x che per y). Ottengo (dopo aver valutato correttamente il numero di cifre significative):

$$\overline{x} = 13.4; s_x = 1.8$$

$$\overline{y} = 35.; s_y = 2.$$

Devo ora valutare se le misure sospette debbano essere rigettate. Trovo il valore di t corrispondente alla distribuzione Normale:

$$t_x = \frac{x_{susp} - \overline{x}}{s_x} = \frac{17 - 13.4}{1.8} = 2$$

$$t_y = \frac{y_{susp} - \overline{y}}{s_y} = \frac{30 - 35.}{2.} = -2.5$$

e valuto il numero n di eventi peggiori del valore sospetto:

$$n(peggiori\ di\ x_{susp}) = N(1 - P(entro\ t_x)) = 8(1 - 0.954) = 0.368 < 1/2$$

$$n(peggiori\ di\ y_{susp}) = N(1 - P(entro\ t_y)) = 8(1 - 0.988) = 0.096 < 1/2$$

Poiche' in entrambi i casi il risultato e' inferiore a 0.5, entrambe le misure di x e y vanno rigettate. Poiche' i dati non sono correlati, avrei potuto rigettare eventualmente uno solo dei valori (x o y).

Nota: In realta' l'approccio piu' rigoroso e' quello di verificare, dopo l'aggiunta del nuovo dato, l'ipotesi "i dati x e y non sono correlati". Nel caso che l'ipotesi vada rigettata, i dati risultano correlati e non si potra' piu' applicare il criterio di Chauvenet (che vale solo per un gruppo di misure della stessa grandezza).

Problema 2

Da una serie di misure di conteggio nell'unita' di tempo, si e' ricavata la seguente distribuzione:

$$k = 0, 1, 2, 3, 4, 5, 6$$
$$n_k = 6, 29, 24, 10, 6, 3, 1$$

Si determinino i parametri della corrispondente:

a) Distribuzione di Poisson,
b) Distribuzione di Gauss
c) Vedere quale delle due distribuzioni approssima meglio i dati.

(N.B. Si ricordi che le distribuzioni di Poisson e di Gauss sono normalizzate...)

Soluzione

Costruisco una tabella con i valori necessari per il calcolo del χ^2. Non utilizzero' i valori per k = 5,6 in quanto i conteggi sono inferiori a 5 e per quanto visto, il test di χ^2 vale per variabili di tipo gaussiano.

k	n_k	$\sqrt{n_k}$	$NP_P(k;\mu)$	$NP_G(k;\mu,\sigma)$
0	6	2.45	11.85	7.90
1	29	5.38	22.12	18.96
2	24	4.90	21.33	24.49
3	10	3.16	13.43	17.38
4	6	2.45	6.32	6.32
5	3			
6	1			
N= 79				

Calcolo i momenti principali della distribuzione:

$$\overline{k} = \frac{\sum_k k\, n_k}{\sum_k n_k} = 1.9\ ;\ s^2 = \frac{1}{N-1}\sum_k (k - \overline{k})^2\, n_k = 1.7$$

Ipotesi Poissoniana: $\mu = \overline{k}$; $\sigma = \sqrt{\overline{k}}$. Calcolo i valori di $NP_P(k;\mu)$ (riportati nella tabella precedente)

$$NP_P(k;\mu) = N\,\frac{\mu^k}{k!}\,e^{-\mu}$$

ed il conseguente χ^2

$$\chi^2 = \sum_{k=0}^{4} \frac{(n_k - N\,P_P(k;\mu))^2}{n_k} = 8.83$$

avendo utilizzato solo cinque misure ed una grandezza calcolata ($\overline{k}$), il numero di gradi di liberta' $\nu = 5 - 1 = 4$ e quindi $\chi^2_\nu = 2.21$.

$$P_4(\chi^2_\nu > 2.21) > 0.05$$

Non posso rigettare l'ipotesi poissoniana.

Ipotesi Gaussiana: $\mu = \overline{k}$; $\sigma = s$. Calcolo i valori di $NP_G(k;\mu,\sigma)$ (riportati nella tabella precedente)

$$NP_G(k;\mu,\sigma) = \frac{N}{\sigma\sqrt{2\pi}} \int_{k-\delta k/2}^{k+\delta k/2} e^{-(k-\overline{k})^2/2\sigma^2}\, dk$$

Una via alternativa che comporta una certa approssimazione (trasformare la gaussiana in un istogramma ad intervalli $\delta k = 1$ centrati in corrispondenza dei valori k), ci permette pero' di semplificare i calcoli. In pratica, l'integrale di cui sopra viene approssimato con

$$NP_G(k;\mu,\sigma) = \frac{N}{\sigma\sqrt{2\pi}} e^{-(k-\overline{k})^2/2\sigma^2}\,\delta k = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,\delta t$$

con $\delta t = \delta k \frac{dt}{dk} = 1 \frac{d}{dk}\frac{k-\overline{k}}{\sigma} = \frac{1}{\sigma}$. Con questa approssimazione si ottengono i risultati della tabella che permettono di calcolare

$$\chi^2 = \sum_{k=0}^{4} \frac{(n_k - N\,P_G(k;\mu,\sigma))^2}{n_k} = 9.55$$

avendo utilizzato solo cinque misure e due grandezze calcolate ($\overline{k}, s$), il numero di gradi di liberta' $\nu = 5 - 2 = 3$ e quindi $\chi^2_\nu = 3.18$.

$$P_3(\chi^2_\nu > 3.18) \simeq 0.02 < 0.05$$

Posso rigettare l'ipotesi gaussiana.

Problema 3

Un'urna contiene 5 palline rosse e 4 bianche. Se estraiamo 3 palline a caso senza rimetterle nell'urna, si determini la probabilita' che:
a) - tutte e tre siano rosse,
b) - tutte e tre siano bianche,
c) - due siano rosse ed una sia bianca,
d) - le palline siano estratte nell'ordine: rossa, bianca, rossa
N.B. vi sono due modi di procedere...

Soluzione

Si tratta di calcolare una probabilita' condizionata, seguendo un ragionamento sequenziale, caso per caso, visto il numero limitato di condizioni. In generale, nei primi tre casi possiamo utilizzare la legge ipergeometrica; mentre nel terzo caso possiamo usare la legge ipergeometrica, ma dividendo il risultato per il numero di combinazioni possibili di 2 palline rosse ed una bianca, in quanto viene richiesta una sequenza ben precisa.

La legge ipergeometrica e' data dalla seguente espressione generale:

$$P(k_1, k_2, ..., k_i; a_1, a_2, ..., a_i) = \frac{\binom{a_1}{k_1}\binom{a_2}{k_2}...\binom{a_i}{k_i}}{\binom{a_1+a_2+...+a_i}{k_1+k_2+...+k_i}}$$

che nel caso dell'esercizio si riduce a

$$P(k_1, k_2; a_1, a_2) = \frac{\binom{a_1}{k_1}\binom{a_2}{k_2}}{\binom{a_1+a_2}{k_1+k_2}}$$

con $a_1 = 5$ rosse, $a_2 = 4$ bianche, $k_1 + k_2 = 3$ estrazioni.
a) $P(3, 0; 5, 4) = \frac{\binom{5}{3}\binom{4}{0}}{\binom{9}{3}} = 0.12$
b) $P(0, 3; 5, 4) = \frac{\binom{5}{0}\binom{4}{3}}{\binom{9}{3}} = 0.05$
c) $P(2, 1; 5, 4) = \frac{\binom{5}{2}\binom{4}{1}}{\binom{9}{3}} = 0.48$
d) $P(2, 1; 5, 4)/C(3, 1) = 0.48/3 = 0.16$.

Applicando la probabilita' condizionata, dobbiamo scrivere:
a) $P(rrr) = P(r)P(r|r)P(r|rr) = (5/9)(4/8)(3/7) = 0.12$
b) $P(bbb) = P(b)P(b|b)P(b|bb) = (4/9)(3/8)(2/7) = 0.05$
calcolo prima "d)" di "c)"...
d) $P(rbr) = P(r)P(b|r)P(r|rb) = (5/9)(4/8)(4/7) = 0.16$
c) $P(rrb) + P(rbr) + P(brr) = 3P(rbr) = 0.48$ (in quanto basta scambiare nel prodotto di frazioni i termini al numeratore, ma il prodotto finale non cambia).