Uso della variabile χ^2: test di verosimiglianza e fit di dati.

C. Meneghini *meneghini@fis.uniroma3.it*

a) La variabile χ^2

La variabile χ^2: se z_1, z_2, ... z_ν sono variabili aleatorie con distribuzione normale standard (N(0,1)) la variabile $\chi^2_\nu = \sum_i z_i^2$ é una variabile aleatoria la cui funzione densitá di probabilitá é:

$$f(x,\nu) = \frac{x^{\nu/2-1}}{2^{\nu/2}\Gamma\left(\nu/2\right)}\, e^{-x/2} \tag{1}$$

Vediamo alcune proprietá della distribuzione della variabile χ^2 che ci saranno utili in seguito:

- il valor medio della distribuzione é eguale al numero di gradi di libertá $\overline{\chi^2_\nu} = \nu$

- la varianza della distribuzione é: $\sigma^2_{\chi^2_\nu} = 2\nu$

- La somma di variabili χ^2 é ancora una variabile χ^2 (proprietá riproduttiva): $\chi^2_{\nu_1} + \chi^2_{\nu_2} = \chi^2_{\nu_1+\nu_2}$

- Per $\nu > 1$ la distribuzione del χ^2 ha un massimo per $\chi = \nu - 2$.

Se X é una variabile aleatoria che segue una distribuzione normale $N(\mu\sigma^2)$ con valore medio μ e varianza σ^2, la variabile normalizzata:

$$Z = \frac{X-\mu}{\sigma}$$

segue una distribuzione normale con media nulla e varianza unitaria (distribuzione normale standard N(0,1)). Quindi se x_1, x_2, ... x_N sono variabili che seguono una distribuzione normale, la variabile:

$$\chi^2_\nu = \sum_{i=1}^{N} \frac{(x_i-\mu)^2}{\sigma^2}$$

é una variabile χ^2 con ν gradi di libertá. Se la variabile χ^2 é costituita da N variabili normali non indipendenti tra loro ma legate da K relazioni matematiche, i gradi di libertá sono $\nu = N - K$

La variabile $\tilde{\chi}^2_\nu$ ridotto: $\tilde{\chi}^2_\nu = \chi^2_\nu/\nu$ é caratterizzata da valor medio $\overline{\tilde{\chi}^2_\nu} = 1$ e varianza $\sigma^2_{\tilde{\chi}^2_\nu} = 2/\nu$

Un a proprietá importante delle variabili χ^2 é che il rapporto tra due variabili χ^2 é ancora una variabile aleatoria che segue la distribuzione F di Fisher:

$$F(\nu_1,\nu_2) = \frac{\chi^2_{\nu_1}/(\nu_1-1)}{\chi^2_{\nu_2}/(\nu_2-1)} = \frac{\tilde{\chi}^2_{\nu_1}}{\tilde{\chi}^2_{\nu_2}}$$

La variabile χ^2 é utilizzata per effettuare test statistici di reiezione di ipotesi e nei metodi di regressione (fitting).

b) Testi del χ^2 per la reiezione delle ipotesi

Un test statistico é una regola di decisione con cui, in base all'osservazione di un campione (dati), si rifiuta o si accetta l'ipotesi H_o riferita alla popolazione intera. Il test del χ^2 rientra nella classe dei cosiddetti **test di significativitá**: sono test che permettono di stabilire (quantificando il margine di rischio) se H_o é incompatibile con i dati sperimentali. Per far questo si definisce una opportuna variabile di test θ con distribuzione data e sene calcola il valore θ_o in base all'ipotesi fatta e al campione osservato. Si calcola quindi la probabilitá che, nell'ipotesi che H_o sia vera, si possa osservare casualmente un valore θ almeno estremo come il valore osservato θ_o: $\alpha_o = P(\theta > \theta_o)$. Se questa probabilitá (che rappresenta il *rischio* di sbagliare) é inferiore ad un limite dato α_L (tipicamente si scelgono valori limite $\alpha_L = 0.05$, $\alpha_L = 0.01$) si rifiuta l'ipotesi H_o, poiché si puó concludere che se H_o fosse vera si sarebbe osservato un evento molto raro (probabilitá minore di α_L), quindi il rischio di sbagliare rifiutando l'ipotesi é piccolo (accettabile).

E' importante sottolineare come i test di reiezione delle ipotesi abbiano lo scopo di quantificare il rischio di sbagliare nel rifiutare l'ipotesi H_o, non il viceversa. Se la probabilitá associata alla variabile di test é piccola: $\alpha_o < \alpha_L$, si puó affermare che *l'ipotesi non é compatibile con i dati con probabilitá* maggiore di $1-\alpha_L$ (confidenza $1 - \alpha_L$, ad esempio per $\alpha_L = 0.05$ si ha una confidenza del 95%). Tuttavia non é vero il contrario: se la probabilitá associata alla variabile di test é alta: $\alpha_o > \alpha_L$, affermare che *l'ipotesi é compatibile con i dati con probabilitá* $1 - \alpha_o$ **é sbagliato**. L'affermazione corretta é: *l'ipotesi non é incompatibile con i dati per il livello di confidenza fissato.*

Il test del χ^2 é utilizzato per stabilire se un modello teorico é incompatibile con un set di osservazioni sperimentali. Per fissare le idee supponiamo di voler confrontare una serie di osservazioni sperimentali (frequenze osservate O_i) con le frequenze attese A_i, calcolate in base a un modello teorico. Se il modello é corretto (ipotesi da sottoporre a test) le osservazioni sperimentali si devono distribuire attorno ai valori attesi A_i con deviazione standard $\sqrt{A_i}$ (nella stima dell'errore abbiamo supposto una distribuzione di Poisson). Costruisco la variabile χ^2:

$$\chi^2_o = \sum_{i=1}^{N} \frac{(O_i - A_i)^2}{A_i},$$

Per sottoporre a test il valore di χ^2_o osservato innanzitutto **i)** di determinano i gradi di libertá del χ^2 calcolato. Questi sono $\nu = N - K$ dove N é il numero di dati sperimentali (numero di variabili z_i che definiscono la χ), K é il numero dei parametri del modello stimati a partire dai dati sperimentali. Ad esempio se utilizzo i dati sperimentali per calcolare media e varianza ($K = 2$) da utilizzare nel modello, il numero di gradi di libertá é: $\nu = N - 2$. Quindi **ii)** si stabilisce un limite α_L alla probabilitá di sbagliare rigettando un'ipotesi corretta (in pratica si stabilisce quale é il rischio che si accetta di correre). Di solito si sceglie $\alpha_L = 0.01$ o $\alpha_L = 0.05$. Il test permette di rigettare l'ipotesi H_o se la probabilitá di osservare un valore di χ^2_ν maggiore del valore trovato: $\alpha_o = P(\chi^2_\nu > \chi^2_o) < \alpha_L$. In tal caso si puó affermare che *la probabilitá che il modello sia inconsistente con i dati*

sperimentali é maggiore di $1-\alpha_L$ (confidenza). Se α_o é alta ($\alpha_o > \alpha_L$) non posso escludere che le differenze trovate siano dovute al caso e quindi il modello **potrebbe** essere corretto. **Attenzione** ribadiamo quanto detto sopra: se la probabilitá associata al χ^2 osservato é maggiore di α_L, l'affermazione: *il modello é corretto con probabilitá* $1-\alpha_L$ é **sbagliata**.

Per valutare il test si possono seguire varie opzioni:

i) una volta deciso il limite di confidenza α_L e determinati i gradi di libertá del problema, si calcola il valore limite $\chi^2_\nu(Lim)$ per il quale sia:

$$\alpha_L = \int_{\chi^2_\nu(Lim)}^{\infty} \frac{x^{\nu/2-1}}{2^{\nu/2}\Gamma\left(\nu/2\right)}\, e^{-x/2} dx$$

Quindi si confronta il valore trovato χ^2_o con $\chi^2_\nu(Lim)$: se $\chi^2_o > \chi^2_\nu(Lim)$ si puó rigettare l'ipotesi con un livello di confidenza maggiore di $1-\alpha_L$. Per il calcolo dei valori limite si possono utilizzare le tabelle disponibili in letteratura (testi di analisi dati, dispense, internet, etc...) oppure si possono calcolare utilizzando un software di analisi statistica. Ad esempio in Excel si utilizza la funzione: **INV.CHI**(α, ν) per il calcolo di $\chi^2_\nu(Lim)$, cioé:

$$\chi^2_\nu(Lim) = \textbf{INV.CHI}(\alpha, \nu)$$

ii) Alternativamente, una volta determinato il numero di gradi di libertá del problema, si calcola la probabilitá associata al valore di una variabile $\chi^2_\nu = \chi^2_o$, cioé la probabilitá di osservare un valore della variabile χ^2_ν maggiore di χ^2_o:

$$\alpha_o = \int_{\chi^2_o}^{\infty} \frac{x^{\nu/2-1}}{2^{\nu/2}\Gamma\left(\nu/2\right)}\, e^{-x/2} dx$$

se $\alpha_o < \alpha_L$ si puó rigettare l'ipotesi. Oppure si puó rigettare l'ipotesi con un rischio α_o. Molti software di analisi statistica prevedono questa opzione. In Excel il calcolo di α_o si ottiene utilizzando la funzione:

$$\alpha_o = \textbf{DISTRIB.CHI}(\chi^2_o, \nu)$$

ESEMPI

Esempio 1) Consideriamo il seguente problema: un costruttore di lampadine afferma che la probabilitá che una delle sue lampadina si fulmini prima di 100 ore di funzionamento é il 10% ($p = 0.1$). Per verificare questa affermazione prendiamo un gruppo di N=90 lampadine e le lasciamo accese per 100 ore. Scaduto il tempo verifichiamo che $O_f = 18$ lampadine sono fulminate e $O_a = 72$ accese. Se il costruttore dicesse il vero ci si dovrebbero avere: $A_f^{th} = Np = 90 \cdot 0.1 = 9$ lampadine fulminate e $A_a^{th} = N(1-p) = 90 \cdot 0.9 = 81$ (valori attesi) lampadine ancora accese. Costruisco la tabella

	Fulminate	Accese	totale
Osservazioni:	17	73	90
Ipotesi:	9	81	90

e quindi il valore di χ^2 associato:

$$\chi^2 = \frac{(18-9)^2}{9} + \frac{(72-81)^2}{81} = 7.9$$

Ho un solo grado di libertá: $\nu = M - K = 2 - 1 = 1$ infatti M=2 sono le osservazioni (numero di lampadine accese e numero di lampadine fulminate) quindi ho usato usato il numero totale di lampadine ($N = O_f + O_a$) per calcolare il numero atteso di lampadine accese (A_a^{th}) o fulminate A_f^{th} in base al modello da testare. Quindi ho utilizzato un grado di libertá dei dati sperimentali per calcolare un solo parametro della distribuzione ipotizzata: K=1.

Il calcolo della probabilitá associata al valore del χ^2 osservato fornisce: $\alpha = DISTRIB.CHI(7.9, 1) = 0.0049 =$ 0.5% il che permette di affermare: *con una grado di confidenza maggiore del 99% l'ipotesi che solo 10% delle lampadine si fulmini prima di 100 ore di funzionamento é falsa.* Oppure: la probabilitá che le lampadine rispettino le specifiche date dal costruttore ma che per caso (grande fiducia nell'onestá del costruttore) in questo esperimento sene sono fulminate di piú é minore di 0.5%

Esempio 2) Si effettuano 18 lanci con una moneta e per 7 volte il risultato é: **testa**. La moneta é ben bilanciata? L'ipotesi da verificare é che la probabilitá che esca testa ad ogni lancio sia $p_T = 0.5$. Si costruisce la seguente tabella:

	T	X	totale
Osservazioni:	7	11	18
Ipotesi:	9	9	18

dove $A_T = A_X = 18 \cdot 0.5$. Quindi il valore di χ^2:

$$\chi^2 = \frac{(7-9)^2}{9} + \frac{(11-9)^2}{9} = 0.89$$

con $\nu = 1$ grado di libertá. La probabiliá associata é: DISTRIB.CHI(0.89,1) = 0.35 = 35%. Dal momento che questa probabilitá é alta non posso escludere che la moneta sia ben bilanciata. Tuttavia **non posso affermare che la moneta sia ben bilanciata.** Infatti sottoponendo a test l'ipotesi che la moneta sia sbilanciata e che la probabilitá di ottenere testa sia solo $p_T = 0.3$, si ottiene la seguente tabella:

	T	X	totale
Osservazioni:	7	11	18
modello:	5.4	12.6	18
$\chi^2 = 0.68$	$\alpha = 0.41$		

da cui risulta che anche in questo caso la probabilitá di osservare un valore di χ^2_1 maggiore di quello osservato é grande ($> 40\%$) quindi non posso escludere neanche l'ipotesi che Testa (T) sia molto sfavorita rispetto a Croce (X). Se si ripete il test per diversi valori di p_T, si vede che, fissando un limite di confidenza $\alpha_L = 0.05$, si puó escludere la consistenza di delle ipotesi con i dati sperimentali con un rischio inferiore a α_L solo per $p_T < 0.2$ o $p_T > 0.6$.

Come fare per restringere il campo di incertezza del test? Si supponga che in un secondo esperimento si siano ottenuti i seguenti risultati: T=70 e X=110. Da un rapido calcolo si ottiene $\chi^2_o = 8.89$ cui é associata una probabilitá $\alpha_o = 0.0029 \ll 0.05$. In questo caso é evidente che l'ipotesi é falsa. Inoltre, facendo un po' di prove, si vede che, fissato il limite di confidenza $\alpha_L = 0.05$, é possibile rigettare le ipotesi se $p_T < 0.32$ o $p > 0.47$. Questo esempio fa capire (sia pure in modo intuitivo) che la numerositá del campione (N_T = numero totale di osservazioni) é rilevante per migliorare la "potenza" del test.

Esempio 3 Distribuzioni discrete: Binomiale.

Per verificare la sensibilitá di una specie di piante ad un virus si espongono al contagio M gruppi (M=90) di $n = 20$ individui ciascuno. Dopo un tempo certo tempo si contano il numero di contagi k per ogni gruppo e si riportano in tabella le frequenze N_k con cui sono osservati k contagi per gruppo. Si vuole verificare l'ipotesi che la probabilitá di contagio sia $p = 0.3$ e che quindi il numero di contagi per gruppo (k) segua una distribuzione binomiale $f(n, p)$.

La distribuzione binomiale calcola la probabilitá di ottenere k di successi (nel nostro caso contagi) su n prove indipendenti (nel nostro caso n=20):

$$f(k,n) = \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!}\, p^k (1-p)^{n-k}$$

In base al modello Binomiale le frequenze teoriche sono:$N_k^{th} = Mf(k, n)$ dove M é il numero delle osservazioni (M=90) e sono riportate in tabella.

Siamo ora in grado di calcolare $\chi^2 = \sum_i z_i^2 = 12.94$. I gradi di libertá sono $\nu = N_{Oss} - K = 11 - 1 = 10$ dove N_{Oss} é il numero di classi k: numero di contagi per gruppo.

In tabella: k = numero di contagi su 20 individui, N_k frequenza osservata, frequenza calcolata in base ad una distribuzione Binomiale $n = 20, p = 0.3$, $z_i^2 = (N_k - N_k^{th})^2/N_k^{th}$. Inoltre $\sum_k N_k = M = 90$, $\sum_k N_k^{th} = M^{th} = 89.2$.

k	N_k	N_k^{th}	z_i^2
2	5	2.506	2.482
3	2	6.444	3.065
4	11	11.738	0.046
5	14	16.098	0.273
6	18	17.248	0.033
7	12	14.784	0.524
8	14	10.296	1.333
9	6	5.883	0.002
10	4	2.774	0.542
11	3	1.081	3.409
12	1	0.347	1.226
tot.	90	89.2	

La probabilitá associata al valore osservato per una variabile di χ^2_{10} con 10 gradi di libertá é:

$$\alpha(12.94, 10) = 22.7\%$$

quindi non si puó rigettare l'ipotesi che il modello sia corretto.

Il calcolo appena effettuato é impreciso, c'é ancora un po' di informazione che non é stata usata e che é bene includere nell'analisi. Infatti non si é tenuto conto del fatto che non ci sono osservazioni per $k < 2$ o $k > 12$: chiamiamo questa *regione esterna* (out), per la quale: $N_{out} = 0$. La somma delle frequenze teoriche $M^{th} = 89.2$ é minore del numero totale di esperimenti M=90. Vuol dire che, in base al modello teorico, ci si aspettano $M - M^{th} = 0.8 = N_{out}^{th}$ osservazioni nella regione esterna. Quindi si puó aggiungere una variabile z al calcolo del χ^2: $z_{out}^2 = (0.8 - 0)^2/0.8 = 0.8$, incrementando di 1 il numero di gradi di libertá. In questo modo si ottiene: $\chi^2 = 13.74$ con $\nu = 11$ gradi di libertá, la cui probabilitá associata é: $\alpha(13.74, 10) = 24.8\%$. Quindi, a maggior ragione non posso scartare la possibiliá che la probabilitá di contagio sia del 30%. O meglio, se lo facessi sbaglierei con una probabilitá elevata $\sim 25\%$!

Nota: se si prova ad effettuare il test per diverse probabilitá di contagio si vede che solo $p < 0.28$ o $p > 0.35$ si

ottengono valori di χ^2 con probabilitá associate inferiori a 0.5%. Si puó quindi escludere che la probabilitá di contagio sia minore di 0.28 o maggiore di 0.35 con una confidenza del 99%. In pratica si é potuto determinato un intervallo di confidenza per la probabilitá di contagio: $p = [0.28; 0.99]$.

Esempio 4 Distribuzioni discrete: distribuzione di Poisson.

Per studiare la distribuzione di piante in una regione si eseguono M=33 campionamenti in zone diverse di eguale superficie Σ. L'istogramma in tabella mostra le frequenze di osservazione. Domanda: in base alle osservazioni si puó assumere che la distribuzione delle piante sia uniforme con densitá $\bar{k}/\Sigma$ e che le osservazioni seguano una distribuzione di Poisson con valore atteso $\lambda = \bar{k}$?

In base ai dati calcoliamo il numero medio di piante per campionamento:

$$\bar{k} = \frac{\sum_i N_k k}{\sum_i N_k} = 28.0$$

e la varianza della distribuzione:

$$\sigma_k^2 = \frac{\sum_i N_k (k - \bar{k})^2}{\sum_i N_k} = 33.8$$

Nel modello Poissoniano la probabilitá di osservare il valore k per un valore atteso λ, é :

$$f(k, \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Utilizziamo come migliore stima di λ il valore medio della distribuzione: $\lambda \sim \bar{k}$ e calcoliamo le frequenze teoriche applicando il modello di Poisson:

$$N_k^{th} = M f(k, \bar{k})$$

In tabella: k = numero di piante osservate per campionamento, frequenza osservata N_k, frequenza calcolata in base al modello di Poisson $\lambda = \bar{k}$, $z_i^2 = (N_k - N_k^{th})^2 / N_k^{th}$, k_{out} rappresenta la regione $k < 20$ e $k > 40$ in cui non ci sono osservazioni ma dove, in base alla distribuzione teorica (esempio precedente), ci si aspettano $N_{out}^{th} = M - \sum_k N_k^{th}$ osservazioni.

k	N_k	N_k^{th}	z_i^2
20	4	0.82	12.430
21	2	1.09	0.763
22	1	1.39	0.108
23	2	1.69	0.057
24	0	1.97	1.975
25	4	2.21	1.441
26	1	2.39	0.806
27	3	2.48	0.110
28	2	2.48	0.093
29	2	2.40	0.066
30	2	2.24	0.026
31	1	2.03	0.519
32	0	1.77	1.774
33	1	1.51	0.171
34	3	1.24	2.486
35	1	1.00	0.000
36	1	0.77	0.065
37	1	0.59	0.291
38	0	0.43	0.433
39	0	0.31	0.311
40	2	0.22	14.562
k_{out}	0	1.98	1.98
tot.	33	33	

Si ottiene $\chi^2 = 40.46$, i gradi di libertá sono: $N_{Oss} - 2 = 22 - 2 = 20$, dove N_{Oss} é il numero delle osservazioni, compresa la regione esterna, 2 sono i parametri della distribuzione teorica calcolati utilizzando i dati sperimentali: il numero totale di osservazioni M e il valor medio $\bar{k}$. Fissando un margine di errore $\alpha_L = 0.05$ si calcola la probabilitá associata al valore osservato di χ^2 con 20 gradi di libertá e si ottiene: $\alpha(40.46, 20) = 0.0049$, il

valore ottenuto é minore di α_L, il che permette di rigettare l'ipotesi che il modello di Poisson sia corretto con un margine di errore del 5%.

Esempio 5 Studio di Tabelle.

Si vuole stabilire se un farmaco efficace. Per questo si sottopone un gruppo di $N_P = 48$ pazienti ad un trattamento con placebo e un gruppo di $N_F = 85$ pazienti ad un trattamento con il farmaco (in totale $N_T = 133$ pazienti) e si ottiene la seguente tabella:

	Efficace	inefficace	totali
Placebo	$N_{PE} = 16$	$N_{Pi} = 32$	$N_P = 48$
Farmaco	$N_{FE} = 40$	$N_{Fi} = 45$	$N_F = 85$
totali	$N_E = 56$	$N_i = 77$	$N_T = 133$

in cui sono riportati il numero totale di casi in cui si é osservato un effetto positivo ($N_E = 56$) indipendentemente dal trattamento, e il numero di casi in cui non si é osservato alcun effetto ($N_i = 77$) indipendentemente dal trattamento.

A prima vista sembra che il farmaco funzioni meglio del placebo: quasi il 50% dei pazienti trattati con il farmaco mostrano un effetto positivo, mentre si osserva un effetto positivo solo nel 30% dei casi trattati con placebo. Vogliamo quantificare la sensazione qualitativa, cioé quantificare il rischio di sbagliare e mettere in vendita un farmaco che casualmente ha funzionato meglio del placebo in questo esperimento. Per dimostrare che il farmaco funziona si fa l'ipotesi che questo non funzioni (ricorda molto il procedimento di dimostrazione per assurdo delle scienze matematiche!) e si verifica se questa ipotesi é inconsistente con i dati, in caso affermativo si puó dire che il farmaco funziona meglio del placebo.

Se il farmaco non funzionasse, quindi se si comportasse come un placebo, le frequenze osservate N_{ij} dovrebbero essere distribuite in modo casuale. Si vuole quindi verificare se le frequenze osservate sono incompatibili con una distribuzione casuale. Essendo N_T il numero totale di pazienti la probabilitá che, indipendentemente da trattamento, uno di essi guarisca é: $f_E = N_E/N_T$ mentre la probabilitá che non ci sia alcun effetto é: $f_I = N_I/N_T$. Inoltre la probabilitá che un paziente sia trattato con il farmaco é: $f_F = N_F/N_T$ mentre la probabilitá che sia trattato con un placebo é: $f_P = N_P/N_T$.

Nell'ipotesi di distribuzione casuale degli effetti la probabilitá che un paziente sia trattato con il placebo e guarisca é $f_{PE} = F_P f_E$, la probabilitá che sia trattato con il farmaco e guarisca é $f_{FE} = f_F f_E$. etc... Quindi si puó costruire la tabella delle frequenze teoriche: $N^{th}_{PE} = N_T f_{PE}$ = numero di successi atteso per un paziente trattato con il placebo, $N^{th}_{FE} = N_T f_{FE}$ = numero di successi atteso per un paziente trattato con il farmaco etc...

	Efficace	inefficace	totali
Placebo	$N^{th}_{PE} = 20.2$	$N^{th}_{Pi} = 27.8$	$N_P = 48$
Farmaco	$N^{th}_{FE} = 35.8$	$N^{th}_{Fi} = 49.2$	$N_F = 85$
totali	$N_E = 56$	$N_i = 77$	$N_T = 135$

Si calcola il valore del χ^2:

$$\chi^2 = \sum_{ij} \frac{(N^{th}_{ij} - N^{exp}_{ij})^2}{N^{th}_{ij}} = \frac{(20.2 - 16)^2}{20.2} + \frac{(35.8 - 40)^2}{35.8} + ... = 2.37$$

da confrontare con la statistica di una variabile χ^2_ν dove il numero di gradi di libertá é $\nu = (n_c - 1)(n_r - 1)$ con n_c: numero di righe e n_r: numero di colonne della tabella. In questo caso $n_r = 2$ e $n_c = 2$ quindi $\nu = 1$. Si ottiene: $\alpha(2.37, 1) = 0.12$ che porta ad escludere l'efficacia del farmaco con un livello di confidenza del 95%. La probabilitá che il farmaco e il placebo non abbiano effetti distinti e che io abbia osservato per caso il valore di $\chi^2 = 2.37$ é maggiore del 10%. Nota: se accettassi un rischio di sbagliare piú alto (es. $\alpha_L = 15\%$) potrei affermare che il farmaco funziona meglio del placebo!!!

Si puó verificare se la tabelle dai dati fosse stata la seguante:

	Efficace	inefficace	totali
Placebo	$N_{PE} = 160$	$N_{Pi} = 320$	$N_P = 480$
Farmaco	$N_{FE} = 400$	$N_{Fi} = 450$	$N_F = 850$
totali	$N_E = 560$	$N_i = 770$	$N_T = 1330$

in cui non cambiano le frequenze relative ma solo il numero totaleN_T, si sarebbe ottenuto: $\chi^2 = 23.71$ la cui probabilitá associata: $\alpha(23.71, 1) = 1.1 \cdot 10^{-6}$ permette di stabilire con buona certezza che il farmaco ha un effetto migliore che del placebo.

c) Criterio di massima verosimiglianza e affinamento dei parametri di una funzione.

Consideriamo variabile aleatoria X che si distribuisce con una funzione di densitá di probabilitá $f(x)$, quindi la probabilitá di ottenere un valore di X compreso tra x e $x + dx$ é $P(x) = f(x)dx$.

La probabilitá di osservare un determinato set di valori x_i (i=1,..N) effettuando N misure indipendenti é il prodotto delle probabilitá di osservare ogni singolo valore:

$$P(x_1; x_2; ...x_N) = \prod_{i=1}^{N} f(x_i)dx = \Lambda dx$$

dove Λ é detta funzione di verosimiglianza (*likelihood*). Per semplicitá consideriamo un set di valori che si distribuiscono con legge normale caratterizzata da valore medio μ_v e varianza σ_v: $N(\mu_v, \sigma_v)$. La probabilitá di osservare un valore x_i é quindi:

$$P(x_i) = f(x_i)dx = \frac{1}{\sqrt{2\pi\sigma_v^2}}\, e^{-\frac{(x_i-\mu_v)^2}{2\sigma_v^2}} dx$$

Per un set di N misure indipendenti x_i $(i = 1, 2, ...N)$ la funzione di verosimiglianza é:

$$\Lambda = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_v^2}}\, e^{-\frac{(x_i-\mu_v)^2}{2\sigma_v^2}}$$

In base ai dati osservati si vuole ottenere la migliore stima dei parametri veri della distribuzione μ_s e σ_s. E' lecito aspettarsi (a posteriori) che l'insieme dei valori effettivamente osservati x_i $(i = 1, 2, ..., N)$ sia quello piú probabile, quindi la migliore stima che posso ottenere per il valore atteso μ_s e la varianza σ_s^2 sono quei valori e che, in base ai dati osservati, rendono massima la probabilitá $P(x_1; x_2; ...x_N)$ cioé la funzione di verosimiglianza. Questo é detto: *criterio della massima verosimiglianza* (*maximum likelihood criterion*). Per calcolare il massimo di una funzione $f(x_1, x_2, ..x_N)$ rispetto ad una delle variabili si cercano gli zeri della derivata prima rispetto a quel parametro: $\delta f/\delta x_i = 0$ (ovviamente c'é anche la condizione $\delta^2 f/\delta x_i^2 < 0$ ma assumiamo sia verificata). Il valore di μ che rende massima Λ é tale che:

$$\left.\frac{\partial\Lambda(\mu,\sigma)}{\partial\mu}\right|_{\mu=\mu_s} = 0$$

Invece della funzione Λ conviene utilizzare il suo logaritmo naturale, la funzione logaritmica ha infatti il pregio di trasformare le produttorie in sommatorie, piú semplici da trattare:

$$\log(\Lambda) = \sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right) = -\sum_{i=1}^{N} \log\left(\sqrt{2\pi\sigma^2}\right) - \sum_{i=1}^{N} \frac{(x_i-\mu)^2}{2\sigma^2}$$

Quindi trovare il massimo di Λ equivale a trovare il minimo di $\log(\Lambda)$, ovvero, dato che σ non dipende da μ:

$$\left.\frac{\partial\Lambda}{\partial\mu}\right|_{\mu=\mu_s} = 0 \quad \text{equivale a :} \quad \left.\frac{\partial}{\partial\mu}\left(\sum_{\mathrm{i}=1}^{\mathrm{N}} \frac{(\mathrm{x_i}-\mu)^2}{\sigma^2}\right)\right|_{\mu=\mu_\mathrm{s}} = 0$$

ma l'ultima sommatoria non é altro che una variabile χ^2 calcolata per l'insieme delle osservazioni. E' abbastanza facile verificare che la derivata della funzione χ^2 rispetto a μ si annulla per:

$$\mu_s = \frac{1}{N}\sum_i x_i$$

quindi la media di N valori é proprio la migliore stima che posso avere del valore atteso μ_v della distribuzione di X. (Nota: si segue un ragionamento analogo per il calcolo della varianza).

Si é visto come il criterio della massima verosimiglianza possa essere usato per stimare i parametri (μ_s, σ_s) di una distribuzione Normale. In effetti possiamo estendere l'applicazione senza modifiche sostanziali ad altre distribuzioni.

Consideriamo un insieme di osservazioni sperimentali Y_i^{exp} $(i = 1, 2, ..N)$ che si suppone seguano una distribuzione di probabilitá $f(p_1, p_2, ...p_m)$ caratterizzata da m parametri. Nell'ipotesi che la distribuzione f sia corretta, i valori osservati devono essere eguali al valore teorico piú un errore statistico:

$$Y_i^{exp} = f_i(p_1, p_2, ...p_m) + \epsilon_i$$

L'errore statistico ha, per definizione, media nulla (della varianza degli errori σ_i^2 ne parleremo piú avanti) e quindi la variabile:

$$z_i = \frac{Y_i^{exp} - f_i(p_1, p_2, ...p_m)}{\sigma_i^2}$$

segue una distribuzione Normale Standard con media nulla e varianza unitaria N(0,1). In base ai dati sperimentali e al modello teorico la funzione di verosimiglianza é:

$$\Lambda(p_1, p_2, ...p_m) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}}\, e^{-\frac{Y_i^{exp} - f_i)^2}{2\sigma_i^2}}$$

funzione di m parametri. Da un punto di vista statistico le migliori stime dei parametri p_j che possiamo ottenere in base ai dati sperimentali sono quelle che rendono massima la funzione di verosimiglianza. Se la varianza degli errori é indipendente dai parametri p_k (questo é un punto da tenere a mente per la discussione che seguirá!), massimizzare la funzione di verosimiglianza equivale (per quanto visto sopra) a cercare il minimo della funzione $\chi^2(p_1, p_2, ...p_m)$:

$$Min\left(\sum_{i=1}^{N} \frac{\left(Y_i^{exp} - f_i(p_1, p_2, ..p_m)\right)^2}{\sigma_i^2}\right)$$

Il metodo del minimo χ^2 si estende naturalmente al raffinamento o fitting, o regressione, etc...) di dati sperimentali. Supponiamo di voler determinare l'accelerazione di gravitá dalla misura del moto di un grave in caduta libera. Per questo abbiamo misurato le posizioni del corpo Y^{exp} in funzione del tempo t_i^{exp}, abbiamo quindi un set di punti nel piano t-Y individuati dalle coordinate (Y_i^{exp}, t_i) $(i = 1, 2, ...N)$

La posizione del corpo in funzione del tempo per un moto rettilineo uniforme é determinata dalla legge:

$$Y^{th}(t) = y_o + v_o(t - t_o) + \frac{1}{2}g(t - t_o)^2 = f(t; y_o, t_o, g)$$

la variabile dipendente Y é funzione della variabile indipendente t e da tre parametri: posizione iniziale y_o, velocitá iniziale v_o e accelerazione di gravitá g.

Ogni misura sperimentale é affetta da un errore statistico (supponiamo ottimisticamente che non ci siano errori sistematici) quindi, assumendo valida le legge del moto, per ogni punto sperimentale si puó scrivere:

$$Y_i^{exp} = Y_i^{th}(t, y_o, v_o, g) + \epsilon_i$$

Ragionando come sopra si puó definire la funzione di verosimiglianza:

$$\Lambda(p_1, p_2, ...p_m) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}}\, e^{-\frac{(Y_i^{exp} - Y_i^{th})^2}{2\sigma_i^2}}$$

e quindi calcolare i parametri della funzione teorica v_o, y_o e g cercando il set di valori che rende massima Λ. Il ragionamento puó essere esteso a qualunque relazione funzionale $Y = f(x_1, x_2, ...x_k; p_1, p_2, ...p_m)$ di k variabili indipendenti e di m parametri. I valori σ_i^2 rappresentano la varianza l'errore associato all'i-esimo punto sperimentale (Y_i^{exp}, t_i). Se σ_i^2 non dipendono dai parametri p_j allora cercare il massimo di Λ in funzione di p_k equivale a cercare il minimo della funzione $\chi^2(x_1, x_2, ...x_k; p_1, p_2, ...p_m)$, come abbiamo visto sopra. La ricerca del minimo della funzione χ^2 é spesso un problema complesso trattato con metodi piú o meno complessi (not linear regression (NLR) methods)

. In generale questi metodi cercano un minimo della funzione $\chi^2(p_1, p_2, ...p_m)$ (o Λ) variando secondo un algoritmo opportuno (es: *Monte Carlo*, *simulated annealing*, *simplex*, *Levenberg-Marquardt*, etc...) i parametri p_j.

Un caso semplice si ha quando:

a la varianza dell'errore eguale per ognuno dei punti sperimentali Y_i^{exp}: $\sigma_i^2 = \sigma^2 = \text{cost}$;

b trascurare gli errori sulle variabili indipendenti $\sigma_{x_i}^2 = 0$

Se sono verificate le condizioni di cui sopra é sufficiente cercare il minimo della funzione:

$$Min\left(\sum_{i=1}^{N}(Y_i^{exp} - f(x; p_1, p_2, ...p_m))^2\right)$$

Questo approccio é noto come *metodo dei minimi quadrati* (*least squares method*). Se la $f(p_1, p_2, ...p_m)$ é un polinomio:

$$f(x; p_o, p_2, ...p_m) = p_m x^m + p_{m-1}x^{m-1} + ... + p_o$$

trovare il set di parametri per cui le condizioni $\partial\chi^2/\partial p_i = (i = 1, 2, ...N)$ sono simultaneamente verificate si riduce ad un problema algebrico relativamente semplice, spesso implementato in molti software di analisi. Se addirittura $f(x; p)$ é una funzione lineare:

$$f(x; a, b) = ax + b$$

siamo nel caso della "regressione lineare semplice" (o regressione lineare multipla se abbiamo k variabili indipendenti) e la funzione $ax + b$ é detta retta di regressione.

La cosa fondamentale da tenere a mente nell'applicare il metodo dei minimi quadrati sono le assunzioni di cui sopra: $\sigma_i^2 = cost$ e $\sigma_{x_i}^2 = 0$. In particolare la condizione $\sigma_{x_i}^2 = 0$ viene spesso dimenticata ma essa é all'origine di un fenomeno apparentemente assai strano: da un punto di vista algebrico, se é valida la relazione lineare $y = f(x) = a_o x + b_o$ che esprime y in funzione di x, allora lo sará anche e la sua inversa $x = f^{-1}(y) = \alpha_o y + \beta_o$, che esprime x in funzione di y con: $a_o = 1/\alpha_o$ e $= b_o - \beta_o/\alpha_o$. Se calcolano le rette di regressione su un set di dati (Y^{exp}, X^{exp}) una volta per $Y = aX + b$ e una volta per $X = \alpha Y + \beta$, le rette trovate saranno, in genere, diverse con $a \neq 1/\alpha$ e $b \neq -\beta/\alpha$. Il motivo é che nel calcolo delle rette di regressione si assume implicitamente nullo l'errore sulle ascisse, quindi nel primo caso $(y = f(x))$ si assume $\sigma_x^2 = 0$, nel secondo caso: $(x = f^{-1}(x))$ si assume $\sigma_y^2 = 0$.

Nel caso in cui la varianza degli errori non sia costante ma si possa ancora trascurare l'errore sulle variabili indipendenti é necessario ricorrere a metodi che minimizzano la funzione χ^2 (a volte detti dei minimi quadrati pesati). Se gli errori non dipendono dai parametri p e si possono trascurare gli errori sulle variabili indipendenti le soluzioni possono essere ottenute algebricamente per funzioni polinomiali e in alcuni casi in cui esse possano essere ricondotte a funzioni polinomiali (*linearizzazione*).

Se le funzioni non sono *linearizzabili*, cioé non possono essere ricondotte a funzioni lineari (o polinomiali) é necessario utilizzare metodi e programmi piú complessi come i metodi di fit non lineare citati sopra.

E' bene tenere a mente il fatto che se non é possibile trascurare gli errori sulle variabili indipendenti non solo il calcolo delle σ_i^2 diventa complicato ma anche la funzione da minimizzare deve essere scelta in modo opportuno. Infatti σ_i^2 puó essere considerato come la varianza della differenza: $(Y_i^{exp} - Y_i^{th}$ per l'i-esimo punto sperimentale, quindi l'errore é una combinazione dell'errore su Y_i^{exp} e dell'errore sul valore atteso Y_i^{th}:

$$\sigma_i^2 = \sigma_{Y_i^{exp}}^2 + \sigma_{Y_i^{th}}^2$$

Usando le regole per la propagazione degli errori casuali si ha:

$$\sigma_{Y_i^{th}}^2 = \left(\frac{\partial f(x; p_1, p_2, ...p_m}{\partial x} \right)^2 \Bigg|_{x=x_i} \sigma_x^2$$

$\sigma_{Y_i^{th}}^2$ é una funzione de parametri p_j. In tal caso trovare il massimo della Λ non é equivalente a trovare il minimo di una funzione χ^2 ma la funzione corretta da minimizzare é:

$$Min \left(\sum_i \left(\log \sigma_i + \frac{(Y_i^{exp} - Y_i^{th})^2}{2\sigma_i^2} \right) \right)$$

... e le cose si complicano.

Errori sui parametri

Una volta trovati i valori dei parametri $p_1^{min}, p_2^{min}, ...p_m^{min}$ che determinano un minimo della funzione χ^2, quale é l'errore associato ad essi? Quale affidabilitá dare ai parametri in termini di *confidenza*?

Una volta trovato il minimo della funzione χ^2_{min} per una determinata scelta dei parametri, si sceglie un valore limite C tale che la probablitá associata a valori del $\chi^2_\nu = \chi^2_{min} + C$ sia piccola (es $\alpha = P(\chi^2_\nu > \chi^2_{min} + C) \leq 5\%$). Quindi si cercano valori estremi dei parametri $p_i^+ = p_i^{min} + \delta p_i$ e $p_i^- = p_i^{min} - \delta p_i$ tali che determinano incrementi del χ^2 rispetto al minimo inferiori a C. I valori δp_i rappresentano gli errori sui parametri raffinati.

Nota: nell'applicare il test del χ^2 lo scopo era verificare l'incompatibilitá di un modello con i dati osservati. Il raffinamento dei parametri di una funzione ha lo scopo di determinare i valori dei parametri che meglio si accordano con i dati sperimentali e valutarne gli errori (intervalli di confidenza). Stabilire il numero corretto di gradi di libertá e valutare bene gli errori σ_i^2 sono due passi chiave per ottenere un buon raffinamento e, soprattutto, per una stima ragionevole degli errori associati ai parametri della funzione teorica.

Il numero di gradi di libertá é $\nu = N - K$ dove K é il numero di parametri liberi.

La stima degli errori ricopre un ruolo essenziale: se il modello é corretto sottostimare gli errori σ_i^2 produce valori di χ^2_{Oss} elevati ($\chi^2_{Oss} \gg \nu - 2$ ovvero $\tilde{\chi}^2_{Oss} \gg 1$), se confrontati con la probabilitá associata ad una variabile χ^2_ν con ν gradi di libertá si puó ottenere una probabilitá molto bassa. Al contrario se gli errori statistici sono sovrastimati si ottengono valori χ^2_o piccoli ($\chi^2_{Oss} \ll \nu - 2$ ovvero $\tilde{\chi}^2_{Oss} \ll 1$). In questo caso la probabilitá associata ad un χ^2_ν é apparentemente molto elevata. In ambedue i casi gli errori stimati su parametri ottenuti potranno essere molto lontani dal vero.

Un criterio per stabilire se la scelta della funzione teorica é corretta e se gli errori sono stati valutati bene é che sia $\chi^2_{Oss} \sim \nu$ ovvero $\chi^2_{Oss}/\nu \sim 1$. In tal caso possiamo avere una certa confidenza nei parametri ottenuti e negli errori calcolati.

Quando un modello é meglio di un altro? Supponiamo di aver ottenuto un buon fit dei dati sperimentali utilizzando diversi modelli, siano m_h parametri utilizzati per il modello h per il quale la funzione di verosimiglianza é Λ_h. Dato che lo scopo é trovare a "massima verosomiglianza" (o minimo χ^2) si potrebbe pensare che un fit é meglio di un altro se Λ é piú grande o se il χ^2 é piú piccolo. Questo non é vero, infatti non solo c'é un limite al numero di parametri che si possono usare ($K < N$) ma aumentando il numero di parametri del fit si rischia di riprodurre artifatti dovuti al rumore piuttosto che alla relazione fisica tra variabili dipendenti e indipendenti.

Come stabilire se un modello é meglio di un altro in base ai dati sperimentali? Nel confrontare due modelli é essenziale tenere conto del numero dei gradi di libertá della funzione χ^2_ν. Se due modelli hanno lo stesso numero di gradi di libertá, é da preferire quello che ha il χ^2 minore. Ma se i gradi di libertá sono diversi bisogna fare attenzione. Infatti aumentando il numero dei parametri liberi generalmente migliora l'accordo, tuttavia, quanto é significativo questo miglioramento. Una valutazione quantitativa sfrutta la statistica della funzione F.

Sappiamo che la differenza di due variabili χ^2 é ancora una variabile χ^2:

$$\chi^2_{\nu_1} - \chi^2_{\nu_2} = \chi^2_{\nu_2-\nu_1}$$

i cui gradi di libetá sono la differenza dei gradi di libetá delle funzioni originarie. Inoltre il rapporto tra due variabili χ^2 é una variabile aleatoria F, quindi:

$$F(\nu_2 - \nu_1, \nu_2) = \frac{(\chi^2_{\nu_1} - \chi^2_{\nu_2})/(\nu_2 - \nu_1)}{\chi^2_{\nu_1}/\nu_2}$$

dove assumiamo $\nu_2 > \nu_1$ e $\chi^2_{\nu_1} > \chi^2_{\nu_2}$ (il modello 1 ha meno gradi di libertá e un χ^2 maggiore). Si confronta quindi la variabile F cosi ottenuta con la statistica della funzione F e si rifiuta il modello piú complesso se la probabilitá associata al valore di F ottenuto é piccola.

Il criterio di Akaike (AIC: Akaike Information Criterion) é un criterio qualitativo per la scelta di un modello. Siano m_h parametri utilizzati per il modello h per il quale la funzione di verosimiglianza é Λ_h.

$$AIC_h = 2m_h - \log \Lambda_h$$

il modello da preferirsi é quello con il valore massimo dell'AIC. Si puó utilizzare la funzione χ^2:

$$AIC_h = 2m_h + n \log \frac{(\chi^2)_h}{N}$$

In questo caso é da preferire il modello con il minimo valore di AIC. Esistono anche formule piú complicate derivate dall'AIC che tengono conto, as esempio, delle dimensioni del campione (piccoli N) e altri efffetti, ma essendo un criterio qualitativo la formula data é sufficient per farsi un idea. Per una determinazione quantitative é meglio utilizzare l'F test descritto sopra.