

Manuale dell'ingegnere intrippato con la statistica

What are the odds?

12 settembre 2019

1 Statistica descrittiva

1.1 Le grandezze che sintetizzano i dati

1.1.1 Media

Dato un insieme x_1, x_2, \dots, x_n di dati, si dice media campionaria la media aritmetica di questi valori.

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

1.1.2 Mediana

Dato un insieme di dati di ampiezza n , lo si ordina dal minore al maggiore. La mediana è il valore che occupa la posizione $\frac{n+1}{2}$ in caso di un insieme dispari, o la media tra $\frac{n}{2}$ e $\frac{n}{2} + 1$ se pari.

1.1.3 Moda

La moda campionaria di un insieme di dati, se esiste, è l'unico valore che ha frequenza massima.

1.1.4 Varianza e deviazione standard campionarie

Dato un insieme di dati x_1, x_2, \dots, x_n , si dice varianza campionaria (s^2), la quantità

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

Una comodità per il calcolo è che

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

Si dice **deviazione standard campionaria** e si denota con s , la quantità

$$s := \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

(la radice quadrata di s^2)

1.1.5 Percentili campionari e box plot

Sia k un numero intero $0 \leq k \leq 100$. Dato un campione di dati, esiste sempre un dato che è contemporaneamente maggiore del k percento dei dati, e minore del $100 - k$ percento. Per trovare questo dato, dati n e $p = \frac{k}{100}$:

1. Disponiamo i dati in ordine crescente
2. Calcoliamo np
3. Il numero cercato è quello in posizione np , arrotondato per eccesso se non intero.

Il 25-esimo percentile si dice *primo quartile*, il 50-esimo *secondo* (ed è pari alla mediana), il 75-esimo *terzo*. Il box plot è un grafica con un quadrato sulla linea dei dati, con i lati sul primo e terzo quartile, e un segno sul secondo.

1.2 Disuguaglianza di Chebyshev

Siano \bar{x} e s media e deviazione standard campionarie di un insieme di dati. Nell'ipotesi che $s > 0$, la disuguaglianza di Chebyshev afferma che per ogni reale $k \geq 1$, almeno una frazione $(1 - 1/k^2)$ dei dati cade nell'intervallo che va da $\bar{x} - ks$ a $\bar{x} + ks$. Usando il ~~pessimo~~ *fantastico* linguaggio da statista: sia assegnato un insieme di dati x_1, \dots, x_n con media campionaria \bar{x} e deviazione standard campionaria $s > 0$. Denotiamo con S_k l'insieme degli indici corrispondenti a dati compresi tra $\bar{x} - ks$ e $\bar{x} + ks$. Sia $\#S_k$ il numero dei suddetti. Allora abbiamo che

$$\frac{\#S_k}{n} \geq 1 - \frac{n-1}{nk^2} > 1 - \frac{1}{k^2}$$

1.3 Insiemi di dati bivariati e coefficiente di correlazione campionaria

A volte non abbiamo a che fare con dati singoli, ma con coppie di numeri, tra i quali sospettiamo l'esistenza di relazioni. Dati di questa forma prendono il nome di *campione bivariato*. Uno strumento utile è il diagramma di dispersione. Una questione interessante è capire se vi sia correlazione tra i dati accoppiati. Parleremo di correlazione positiva quando abbiamo una proporzionalità diretta tra i due, di correlazione negativa quando abbiamo una proporzionalità inversa.

1.3.1 Coefficiente di correlazione campionaria

Dato un campione bivariato (x_i, y_i) , sono definite le medie \bar{x} e \bar{y} . Possiamo senz'altro dire che se un valore x_i è grande rispetto alla media, la differenza $x_i - \bar{x}$ sarà positiva, mentre se x_i è piccolo, la differenza sarà negativa. Quindi, considerando il prodotto $(x_i - \bar{x})(y_i - \bar{y})$, sarà positivo per correlazioni positive, negativo per correlazioni negative. Se l'intero campione mostra quindi un'elevata correlazione, ci aspettiamo che la somma di tutti i prodotti $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ darà una buona stima della correlazione. Normalizziamola dividendo per $(n-1)$ e per il prodotto delle deviazioni standard campionarie, e otteniamo il **coefficiente di correlazione campionaria**

$$r := \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$$

con s_x e s_y deviazioni standard campionarie di x e y .

1.3.2 Proprietà del coefficiente di correlazione campionaria

Sebbene parleremo meglio di questo bastardo nella sezione sulla regressione, elenchiamo qui alcune proprietà:

1. $-1 \leq r \leq 1$
2. Se per opportune costanti a e b , con $b > 0$ sussiste la relazione lineare $y_i = a + b x_i$, allora $r = 1$.
3. Se per opportune costanti a e b , con $b < 0$ sussiste la relazione lineare $y_i = a + b x_i$, allora $r = -1$.
4. Se r è il coefficiente di correlazione del campione (x_i, y_i) , $i = 1, \dots, n$, allora lo è anche per il campione $(a + b x_i, c + d y_i)$, purché le costanti a e b abbiano lo stesso segno.

2 Elementi di probabilità

2.1 Spazio degli esiti ed eventi

Si dice spazio degli esiti l'insieme di tutti gli esiti possibili di un esperimento. Se ad esempio l'esito dell'esperimento fosse il sesso di un neonato, lo spazio degli esiti sarebbe

$$S = \{f, m\}$$

I sottoinsiemi dello spazio degli esiti si dicono **eventi**, quindi un evento E è un insieme i cui elementi sono esiti possibili. Si dice E^c l'opposto dell'evento, quindi $P(E^c) = 1 - P(E)$. Risulta ovvio che $1 = P(E^c) + P(E)$. Se abbiamo due eventi qualsiasi, la loro unione $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

2.2 Spazi di esiti equiprobabili

Per tanti esperimenti è naturale assumere che ognuno degli esiti abbia la stessa probabilità di accadere. Abbiamo quindi che la probabilità che E accada è pari a $P(E) = \frac{1}{N}$.

2.2.1 Principio di enumerazione

Consideriamo la realizzazione di due diversi esperimenti che possono avere rispettivamente m ed n esiti. Allora complessivamente avremo mn risultati.

2.3 Coefficiente binomiale

Vogliamo ora determinare il numero di diversi gruppi di r oggetti che si possono formare scegliendoli da un insieme di n . Ad esempio, quanti gruppi di 3 lettere possono formarsi dal gruppo $\{A, B, C, D, E\}$. In generale, poiché il numero di modi diversi di scegliere r oggetti su n tenendo conto dell'ordine è dato da $n(n-1)\dots(n-r+1)$, e poiché ogni gruppo di lettere viene contato $r!$ volte (uno per permutazione), il numero di gruppi di r elementi su n totali è dato da

$$\frac{n(n-1)\dots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!} = \binom{n}{r}$$

2.4 Probabilità condizionata

Vogliamo ora calcolare la probabilità che un evento accada, appurato che ne è accaduto un altro. Ad esempio, lanciamo due dadi. L'evento E cercato è che il risultato sia 8. L'evento F già accaduto è che il primo dato risulta in un 3. Si dice probabilità condizionata di E dato F

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

2.5 Fattorizzazione di un evento e formula di Bayes

Siano E ed F due eventi qualsiasi. È possibile esprimere E come

$$P(E) = P(E \cap F) + P(E \cap F^c)$$

Visto inoltre che i due sono eventi disgiunti, si ha che

$$P(E) = P(E|F)P(F) + P(E|F^c)P(F^c)$$
$$P(E|F)P(F) + P(E|F^c)[1 - P(F)]$$

Questa *orribile* equazione, ci mostra che la probabilità dell'evento E si può ricavare come media pesata delle probabilità condizionali di E sapendo che: F si è verificato e non si è verificato. I pesi sono ovviamente le probabilità degli eventi a cui si condiziona.

2.6 Eventi indipendenti

Due eventi si dicono indipendenti quando il risultato di uno non influenza l'altro. In altre parole, significa che avendo due eventi E ed F, se so che F è accaduto, la probabilità che accada E non cambia.

$$P(E \cap F) = P(E)P(F)$$

3 Variabili aleatorie e valore atteso

Quando realizziamo un esperimento casuale, non sempre siamo interessati a tutti i risultati del suddetto. Se ad esempio lanciassimo due dadi, potrebbe interessarci la sola somma e non i singoli risultati. Queste quantità di interesse sono dette **variabili aleatorie**. Siccome il valore di questa variabile è dato dal risultato dell'esperimento, possiamo assegnare delle probabilità a queste. Queste variabili aleatorie hanno una *funzione indicatrice* definita, ad esempio, così:

$$I := \begin{cases} 1 & \text{se } X = 1 \text{ o } 2 \\ 0 & \text{se } X = 0 \end{cases}$$

Variabili aleatorie con un numero finito o numerabile di valori possibili sono dette **discrete**. Esistono anche variabili aleatorie **continue**.

3.0.1 Funzione di ripartizione

La funzione di ripartizione F di una variabile aleatoria X, è definita, per ogni numero reale x, tramite

$$F(x) := P(X \leq x)$$

Quindi $F(x)$ esprime la probabilità che la variabile aleatoria X assuma un valore *minore o uguale* a x. Tutte le questioni di probabilità che si possano sollevare su una variabile aleatoria, ammettono una risposta in termini della sua funzione di ripartizione.

3.1 Variabili aleatorie discrete e continue

Se X è una variabile aleatoria discreta, la sua funzione di massa di probabilità, o funzione di massa, si definisce nel modo seguente:

$$p(a) := P(X = a)$$

La funzione $p(a)$ è non nulla su un insieme al più numerabile di valori. Infatti, se x_1, x_2, \dots, x_n sono i possibili valori di X , allora

$$\begin{aligned} p(x_i) &> 0 \quad i = 1, 2, \dots \\ p(x) &= 0 \quad \text{tutti gli altri valori di } x \end{aligned}$$

Siccome X deve assumere i suddetti valori, necessariamente deve essere vero che

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

Una variabile aleatoria che possa assumere un'infinità non numerabile di valori, non potrà essere discreta. Si dirà **continua** se esiste una funzione non negativa f , definita su tutto \mathbb{R} , avente la proprietà che per ogni insieme B di numeri reali,

$$P(X \in B) = \int_B f(x) dx$$

Questa funzione è detta **funzione di densità di probabilità**. L'equazione dice che la probabilità che una variabile aleatoria continua X appartenga a un insieme B si può trovare integrando la sua densità su tale insieme. Pare ovvio che

$$1 = P(X \in \mathbb{R}) = \int_{-\infty}^{\infty} f(x) dx$$

Tutte le probabilità che riguardano una variabile aleatoria continua possono essere espresse in funzione della sua densità di probabilità:

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

Se poniamo $b = a$, notiamo che la probabilità che una variabile aleatoria continua assuma un valore particolare a è nulla:

$$P(X = a) = \int_a^a f(x) dx = 0$$

Leghiamo la funzione di ripartizione F alla densità f così:

$$F(a) := P(X \in (-\infty, a]) = \int_{-\infty}^a f(x) dx$$

Derivando entrambi otteniamo la relazione fondamentale:

$$\frac{d}{da} F(a) = f(a)$$

La densità è quindi la derivata della funzione di ripartizione. Notiamo che quando conosciamo la funzione di massa di probabilità di una variabile aleatoria discreta, o la funzione di densità di probabilità di una continua, abbiamo abbastanza informazioni per poter calcolare le probabilità di ogni evento che dipenda dalla sola variabile aleatoria.

3.2 Coppie e vettori di variabili aleatorie

Ci sono situazioni in cui abbiamo necessità di studiare le **relazioni** tra variabili aleatorie multiple. Per specificare la relazione tra due variabili aleatorie X e Y , il primo passo è estendere il concetto di funzione di ripartizione. Siano quindi X e Y due variabili aleatorie che riguardano lo stesso esperimento casuale. Si dice *funzione di ripartizione congiunta* di X e Y la funzione di due variabili seguente:

$$F(x, y) := P(X \leq x, Y \leq y)$$

dove la virgola denota l'intersezione tra gli eventi. La conoscenza di questa funzione permette, almeno in teoria, di calcolare le probabilità di tutti gli eventi che dipendono, singolarmente o congiuntamente, da X e Y .

3.2.1 Distribuzione congiunta per variabili aleatorie discrete

Se sappiamo che un vettore aleatorio è di tipo discreto, possiamo definire e utilizzare la funzione di massa di probabilità. Se X e Y sono variabili aleatorie discrete che assumono i valori x_1, x_2, \dots e y_1, y_2, \dots , la funzione

$$p(x_i, y_j) := P(X = x_i, Y = y_j), \quad i = 1, 2, \dots \quad j = 1, 2, \dots$$

è la loro funzione di massa di probabilità congiunta. Le funzioni di massa individuali si possono ricavare da questa, notando che, siccome Y deve assumere uno dei valori y_j , l'evento $\{X = x_i\}$, può essere visto come l'unione al variare di j degli eventi $\{X = x_i, Y = y_j\}$, che sono mutuamente esclusivi. Da qui:

$$p_X(x_i) := P(X = x_i) = \sum_j p(x_i, y_j)$$

Anche se le individuali possono essere ricavate dalla congiunta, la congiunta non può essere ricavata dalle condizionali.

3.2.2 Distribuzione congiunta per variabili aleatorie continue

Due variabili aleatorie X e Y sono congiuntamente continue se esiste una funzione non negativa $f(x, y)$ tale che, per ogni sottoinsieme C del piano cartesiano,

$$P((X, Y) \in C) = \int \int_{(x, y) \in C} f(x, y) dx dy$$

questa è detta **densità congiunta** delle variabili aleatorie X e Y . Otteniamo inoltre che

$$P(X \in A, Y \in B) = \int_B \int_A f(x, y) dx dy$$

E, in conclusione,

$$P(X \in A) = \int_A f_X(x) dx$$

Per ricavare le individuali, otteniamo

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y) dy \\ f_Y(y) &= \int_{-\infty}^{\infty} f(x, y) dx \end{aligned}$$

3.2.3 Variabili aleatorie indipendenti

Due variabili aleatorie sono indipendenti se tutti gli eventi relativi alla prima sono indipendenti da tutti quelli relativi alla seconda. La definizione è che se, per ogni coppia di insiemi di numeri reali A e B è soddisfatta

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

le due V.A. sono indipendenti. Se le V.A. sono discrete, l'equazione equivale a dire che la funzione di massa congiunta è il prodotto delle marginali:

$$p(x, y) = p_X(x)p_Y(y)$$

Possiamo generalizzare le osservazioni suddette anche per vettori di variabili aleatorie. *Lo faremo? Non credo proprio.*

3.2.4 Distribuzioni condizionali

Le relazioni esistenti tra due variabili aleatorie possono essere chiarite dallo studio della distribuzione condizionale di una delle due, dato il valore dell'altra. Si ricorda che che presi comunque due eventi E e F con $P(F) > 0$, la probabilità di E condizionata a F è data da

$$P(E|F) := \frac{P(E \cap F)}{P(F)}$$

è naturale applicare questo schema anche alle variabili aleatorie discrete. Siano X e Y due variabili aleatorie discrete con funzione di massa congiunta $p(\cdot, \cdot)$, diciamo funzione di massa di probabilità condizionata di X dato Y e si indica con $p_{X|Y}(\cdot|\cdot)$, la funzione di due variabili così definita:

$$p_{X|Y}(x|y) := P(X = x|Y = y) \\ \frac{p(x, y)}{p_Y(y)}, \quad \forall x \forall y \text{ con } p_Y(y) > 0$$

Se y non è un valore possibile di Y , ovvero se $P(Y = y) = 0$, la quantità $p_{X|Y}(x|y)$ non è definita.

3.3 Valore atteso

Uno dei concetti più importanti di tutta la teoria della probabilità (*ziocane*) è quello di valore atteso. Esso è definito come il numero

$$E[X] := \sum_i x_i P(X = x_i)$$

In altri termini, si tratta della media pesata dei valori possibili di X , usando come pesi le probabilità che vengano assunti. È ovvio che nel caso di V.A. continue, il giochino non funziona. Definiamo quindi il valore atteso di una V.A. continua con funzione di densità f , come

$$E[X] := \int_{-\infty}^{\infty} x f(x) dx$$

3.4 Proprietà del valore atteso

Consideriamo una V.A. di cui conosciamo la distribuzione. *What if*, se anziché calcolare il valore atteso di X , volessimo calcolare quello di una funzione $g(X)$? Notiamo che $g(X)$ è comunque una variabile aleatoria. Ricaviamo quindi la sua distribuzione, e ne calcoliamo il valore atteso. Ponendo le cose in maniera rigorosa:

$$E[g(X)] = \sum_x g(x)p(x) \text{ per V.A. discrete}$$
$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx \text{ per V.A. continue}$$

Per ogni coppia di costanti reali a e b , abbiamo anche che

$$E[aX + b] = aE[X] + b \quad \text{e quindi} \quad E[aX] = aE[X]$$

3.4.1 Valore atteso della somma di variabili aleatorie

Con *complessi calcoli tendenzialmente inutili*, otteniamo che

$$E[X + Y] = E[X] + E[Y]$$

Tale risultato vale sia nel caso discreto, che in quello continuo.

3.5 Varianza

A volte, conoscere la media di una distribuzione non basta. *Se inserissimo l'autore di questi riassunti con la testa in un freezer e i piedi in un forno, la temperatura media sarebbe abbastanza ok, l'autore no.* Per questo, è utile conoscere quanto i valori si allontanano dalla media. Questo è proprio il compito della **varianza**. Sia X una variabile aleatoria con media μ , la varianza di x , che denotiamo con $Var(X)$ è la quantità

$$Var(X) := E[(X - \mu)^2]$$

o, in alternativa (*questa è molto più comoda*)

$$Var(X) = E[X^2] - E[X]^2$$

3.6 La covarianza e la varianza della somma di V.A.

Come sappiamo, la media della somma di V.A. coincide con la somma delle loro medie. Per la varianza, in generale, questo non è vero. **In un caso sì: quando le V.A. sono indipendenti.** Prima di tutto, però, definiamo il concetto di Covarianza: date due V.A. X e Y di media μ_X e μ_Y , essa vale

$$Cov(X, Y) := E[(X - \mu_X)(Y - \mu_Y)]$$

o, in alternativa

$$Cov(X, Y) := E[XY] - E[X]E[Y]$$

Derivano anche alcune semplici proprietà

$$Cov(X, Y) = Cov(Y, X)$$
$$Cov(X, X) = Var(X)$$
$$Cov(aX, Y) = aCov(X, Y) = Cov(X, aY)$$

E se avessimo 3 V.A.?

$$\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$$

Inoltre, generalizzando i concetti, se avessimo n V.A. X_1, \dots, X_n e n Y_1, \dots, Y_n

$$\text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, Y_j)$$

3.6.1 Variabili aleatorie indipendenti

Se abbiamo due V.A. X e Y indipendenti, sappiamo che

$$E[XY] = E[X]E[Y]$$

Questo implica inoltre che

$$\text{Cov}(X, Y) = 0$$

e quindi, se abbiamo n V.A., la varianza della somma è la somma delle varianze.

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)$$

4 La funzione generatrice dei momenti

La *funzione generatrice dei momenti*, o più semplicemente, la funzione generatrice ϕ di una V.A. X, è definita, per tutti i t reali per i quali il valore atteso di e^{tX} ha senso, dall'espressione

$$\phi(t) := E[e^{tX}] = \begin{cases} \sum_x e^{tx} p(x) & \text{se X è discreta} \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx & \text{se X è continua} \end{cases}$$

Il nome deriva dal fatto che tutti i momenti di cui è dotata X possono essere ottenuti derivando più volte nell'origine la funzione $\phi(t)$. Ad esempio,

$$\phi'(t) = \frac{d}{dt} E[e^{tX}] = E[Xe^{tX}]$$

Quindi, $\phi'(0) = E[X]$, e, più in generale, $\phi^n(0) = E[X^n]$. Se X e Y sono variabili indipendenti con funzioni generatrici ϕ_X e ϕ_Y , e se ϕ_{X+Y} è la funzione generatrice dei momenti di $X + Y$, allora

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$$

Un'osservazione interessante sulla generatrice dei momenti, è che essa *determina la distribuzione*, ossia se due V.A. hanno identica generatrice, hanno identica legge (quindi funzione di ripartizione e funzione di massa).

5 La legge debole dei grandi numeri

Per introdurre la suddetta, prima enunciamo la **disuguaglianza di Markov**: se X è una variabile aleatoria che non è mai negativa. allora per ogni $a > 0$

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Come corollario, ricaviamo la disuguaglianza di Chebyshev: data una V.A. X con media μ e varianza σ^2 , allora per ogni $r > 0$

$$P(|X - \mu| \geq r) \leq \frac{\sigma^2}{r^2}$$

Otteniamo infine la **legge debole dei grandi numeri**. *No, non quella con cui giustificate il vostro provarci con ogni essere vivente femminile.* Sia X_1, X_2, \dots una successione di variabili aleatorie i.i.d (indipendenti, identicamente distribuite), tutte con media μ . Allora, per ogni $\epsilon > 0$:

$$P\left(\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| > \epsilon\right) \rightarrow 0 \quad \text{quando } n \rightarrow \infty$$

Ora, vi chiederete: cosa me ne faccio? Posso spiegarla alle tipe in discoteca? No. Un'applicazione interessante è la seguente: supponiamo di ripetere in successione molte copie indipendenti di un esperimento, in ciascuna delle quali può verificarsi un certo evento E :

$$X_i := \begin{cases} 1 & \text{se } E \text{ si realizza nell'esperimento } i\text{-esimo} \\ 0 & \text{se } E \text{ non si realizza} \end{cases}$$

la sommatoria $X_1 + \dots + X_n$ rappresenta il numero di prove - tra le prime n - in cui si è verificato l'evento E . Poiché

$$E[X_i] = P(X_i = 1) = P(E)$$

si deduce che la frazione delle n prove nelle quali si realizza E , tende alla probabilità $P(E)$.

6 Modelli di variabili aleatorie

*Siamo giunti a una sezione tanto interessante, quanto fastidiosa: quella in cui dovete ricordare uno sh*t ton di roba.*

6.1 Variabili aleatorie di Bernoulli e binomiali

Supponiamo di fare un esperimento che ha solo due esiti, *successo e fallimento*. Sappiamo che

$$\begin{aligned} P(X = 0) &= 1 - p \\ P(X = 1) &= p \end{aligned}$$

Una V.A. con funzione di massa di probabilità come questa, è detta **Bernoulliana**. Il suo valore atteso $E[X] = p$. Supponiamo ora di realizzare n esperimenti, ciascuno dei quali è descritto da una Bernoulliana. Se X denota il numero totale di successi, X si dice V.A. binomiale di parametri (n, p) . La funzione di massa di probabilità è data da:

$$P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i}$$

con il solito coefficiente binomiale

$$\binom{n}{i} := \frac{n!}{i!(n-i)!}$$

Si noti che la somma delle probabilità di tutti i valori possibili è ovviamente 1.

$$\sum_i P(X = i) = [p + (1 - p)]^n = 1$$

Osserviamo che se X_1 e X_2 sono binomiali di parametri (n_1, p) e (n_2, p) e sono indipendenti, allora la somma $X_1 + X_2$ è binomiale di parametri $(n_1 + n_2, p)$.

6.1.1 Calcolo esplicito della distribuzione binomiale

Supponiamo che X sia binomiale di parametri (n, p) . Per poter calcolare operativamente la funzione di ripartizione:

$$P(X \leq i) = \sum_{k=0}^i \binom{n}{k} p^k (1-p)^{n-k}$$

o la funzione di massa:

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$$

è molto utile sapere che

$$P(X = k+1) = \frac{p}{1-p} \frac{n-k}{k+1} P(X = k)$$

6.2 Variabili aleatorie di Poisson

Le variabili aleatorie di Poisson assumono solo valori interi non negativi. In altri termini, definiamo *una variabile aleatoria X che assuma i valori $0, 1, 2, \dots$ poissoniana* di parametro λ , $\lambda > 0$, se la sua funzione di massa di probabilità è data da

$$P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}$$

Calcoliamo la generatrice dei momenti $E[e^{tX}] = e^{-\lambda} e^{\lambda e^t}$, per poi derivarla e calcolarla in $X = 0$. Troviamo quindi che

$$\begin{aligned} E[X] &= \phi'(0) = \lambda \\ \text{Var}(X) &= \phi''(0) - E[X]^2 = \lambda \end{aligned}$$

Una caratteristica interessante della poissoniana è che può essere utilizzata come approssimazione di una binomiale di parametri (n, p) , quando n è molto grande e p molto piccolo. Ponendo $\lambda = np$:

$$P(X = i) \approx \frac{\lambda^i}{i!} e^{-\lambda}$$

In altri termini, il totale dei successi di un numero n molto elevato di ripetizioni dell'esperimento, è una variabile aleatoria di con distribuzione approssimativamente di Poisson, con media $\lambda = np$.

6.2.1 Calcolo esplicito della distribuzione di Poisson

Se X è una variabile aleatoria di Poisson di media λ , allora

$$\frac{P(X = i+1)}{P(X = i)} = \frac{\lambda^{i+1} e^{-\lambda}}{(i+1)! \lambda^i e^{-\lambda}} = \frac{\lambda}{i+1}$$

A cosa serve sto pippozzo? Possiamo usarla per calcolare, a partire da $P(X = 0) = e^{-\lambda}$, tutti gli altri. Ad esempio:

$$\begin{aligned} P(X = 1) &= \lambda P(X = 0) \\ P(X = 2) &= \frac{\lambda}{2} P(X = 1) \end{aligned}$$

6.3 Variabili aleatorie ipergeometriche

Una scatola contiene N batterie accettabili, ed M difettose. Si estraggono senza rimessa n batterie, dando pari probabilità a ciascuno degli $\binom{N+M}{n}$ sottoinsiemi possibili. Se denotiamo con X il numero di batterie accettabili estratte,

$$P(X = i) = \frac{\binom{N}{i} \binom{M}{n-i}}{\binom{N+M}{n}} \quad i = 0, 1, \dots, n$$

Una V.A. con massa di probabilità data da questa equazione, è detta **variabile aleatoria ipergeometrica** di parametri N , M e n .

6.4 Variabili aleatorie uniformi

Una variabile aleatoria continua si dice uniforme sull'intervallo $[\alpha, \beta]$, se ha funzione di densità data da:

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{se } \alpha \leq x \leq \beta \\ 0 & \text{altrimenti} \end{cases}$$

Si noti che soddisfa le condizioni per essere una densità di probabilità, in quanto

$$\int_{-\infty}^{\infty} f(x) dx = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} dx = 1$$

Per poter assumere la distribuzione uniforme, nella pratica, occorre che la V.A. abbia come valori possibili i punti di un intervallo limitato $[\alpha, \beta]$, inoltre si deve poter supporre che essa abbia le stesse probabilità di cadere vicino a un qualunque punto dell'intervallo. La probabilità che una V.A. X , uniforme su $[\alpha, \beta]$ è pari al rapporto tra le lunghezze dei due intervalli. Infatti, se $[a, b]$ è contenuto in $[\alpha, \beta]$

$$P(a < X < b) = \frac{1}{\beta - \alpha} \int_a^b dx = \frac{b - a}{\beta - \alpha}$$

6.5 Variabili aleatorie normali

Questa è la sezione più applicabile alla vita reale. *Quella che, a forza di studiare statistica, avete dimenticato.* Una variabile aleatoria X si dice **normale** oppure **gaussiana** di parametri μ, σ^2 e si scrive $X \sim N(\mu, \sigma^2)$ se X ha funzione di densità data da

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\} \quad \forall x \in \mathbb{R}$$

La densità normale è una curva a campana simmetrica rispetto all'asse $x = \mu$, dove ha il massimo pari a $(\sigma\sqrt{2\pi})^{-1} \approx 0.399/\sigma$. Come sempre, cerchiamo valore atteso e varianza, calcolando la generatrice $\phi = E[e^{tX}]$, derivandola e calcolandola in $X = 0$. Otteniamo

$$\begin{aligned} E[X] &= \phi'(0) = \mu \\ \text{Var}(X) &= E[X^2] - E[X]^2 = \sigma^2 \end{aligned}$$

Un risultato importante è che se X è una gaussiana e Y una trasformazione lineare di X , allora Y è a sua volta una gaussiana. In altri termini: sia $X \sim N(\mu, \sigma^2)$ e sia $Y = \alpha X + \beta$, dove α e β sono due costanti

reali, con $\alpha \neq 0$. Allora Y è una normale con media $\alpha\mu + \beta$ e varianza $\alpha^2\sigma^2$. Un corollario della suddetta permette di ottenere il fantomatico processo di normalizzazione: se $X \sim N(\mu, \sigma^2)$, allora

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

è una V.A. normale con media 0 e varianza 1. Questa roba è così importante che ha un nome, **normale standard**, e un simbolo:

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy \quad \forall x \in \mathbb{R}$$

Il fatto che $Z := (X - \mu)/\sigma$ abbia distribuzione normale standard quando X è gaussiana di media μ e varianza σ^2 , ci permette di esprimere le probabilità relative di X in termini di probabilità su Z . Ad esempio per trovare $P(X < b)$, notiamo che $X < b$ s.s.se

$$\frac{X - \mu}{\sigma} < \frac{b - \mu}{\sigma}$$

così che

$$P(X < b) = \Phi\left(\frac{b - \mu}{\sigma}\right)$$

Analogamente, per ogni $a < b$, si ha che

$$P(a < X < b) := \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

6.6 Variabili aleatorie esponenziali

Una V.A. continua, la cui funzione di densità è data da

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases}$$

per un opportuno valore della costante $\lambda > 0$, si dice **esponenziale** con parametro (*o intensità*) λ . La funzione di ripartizione di una tale variabile è data da

$$F(x) = P(X \leq x) = 1 - e^{-\lambda x} \quad x \geq 0$$

Nella pratica, la distribuzione esponenziale può rappresentare il tempo di attesa prima che si verifichi un certo evento casuale. Calcoliamo la generatrice dei momenti $E[e^{tX}] = \frac{\lambda}{\lambda - t}$ con $t < \lambda$. Derivando e calcolando in 0:

$$E[X] = \phi'(0) = \frac{1}{\lambda}$$

$$Var(X) = E[X^2] - E[X]^2 = \frac{1}{\lambda^2}$$

La proprietà fondamentale dell'esponenziale è la sua **assenza di memoria**. *Che cazzo vuol dire? Nemmeno io ho memoria ma non ho libri da 40 euro che parlano di me.* Immaginiamo che X rappresenti il tempo che passa prima che un oggetto si rompa, dato t il tempo per cui ha già funzionato. Chiaramente, la probabilità che l'oggetto duri ancora per un tempo s è $P(X > s + t | X > t)$. Otterremo che la condizione $X > t$ non varia la probabilità, e non ci dà quindi "informazioni utili". Un'altra informazione utile, è che se abbiamo n V.A. esponenziali X_1, \dots, X_n indipendenti, di parametri $\lambda_1, \dots, \lambda_n$, allora la V.A. $Y := \min(X_1, \dots, X_n)$ è un'esponenziale di parametro $\lambda_Y = \sum_{i=1}^n \lambda_i$.

6.6.1 Il processo di Poisson

Consideriamo una serie di eventi istantanei che avvengono a intervalli di tempo casuali, e sia $N(t)$ il numero di quanti se ne sono verificati nell'intervallo $[0, t]$. $N(t)$ si dice *processo di Poisson* di intensità λ , $\lambda > 0$, se

- $N(0) = 0$: questa stabilisce che si iniziano a contare gli eventi al tempo 0.
- Il numero degli eventi che hanno luogo in intervalli di tempo disgiunti sono indipendenti
- La distribuzione del numero di eventi che si verifica in un dato intervallo di tempo dipende solo dalla lunghezza dell'intervallo, e non dalla sua posizione.
- $\lim_{h \rightarrow 0} \frac{P(N(h)=1)}{h} = \lambda$
- $\lim_{h \rightarrow 0} \frac{P(N(h) \geq 2)}{h} = 0$

Le ultime due affermano che se si considera un intervallo di tempo di lunghezza h , vi è approssimativamente possibilità λh che vi occorra un evento solo, e circa una probabilità nulla che se ne verifichino due o più. Se $N(t)$ è un processo di Poisson di intensità λ , allora

$$P(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

Ovvero, il numero di eventi in $[0, t]$ ha distribuzione di Poisson di media λt . Inoltre, i tempi che separano gli eventi di un processo di Poisson, sono una successione di V.A. esponenziali di intensità λ .

6.7 Variabili aleatorie di tipo gamma

Una variabile aleatoria continua si dice avere distribuzione di tipo gamma di parametri (α, λ) , con $\alpha > 0, \lambda > 0$ se la sua funzione di densità di probabilità è data da

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases}$$

dove Γ denota la funzione gamma di Eulero, che è definita in modo da normalizzare l'integrale di f . Per proprietà che ci asterremo dall'enunciare, possiamo calcolare la gamma sugli interi così:

$$\Gamma(n) = (n-1)!$$

Si noti che per $\alpha = 1$ la distribuzione gamma coincide con l'esponenziale. Troviamo ora la generatrice dei momenti, la deriviamo e calcoliamo in 0 per ottenere:

$$E[X] = \phi'(0) = \frac{\alpha}{\lambda}$$
$$Var(X) = E[X^2] - E[X]^2 = \frac{\alpha}{\lambda^2}$$

Se abbiamo più V.A. indipendenti di tipo gamma X_i , allora $\sum_i^n X_i$ è una gamma di parametri $(\sum_i^n \alpha_i, \lambda)$. Di conseguenza, se abbiamo una somma di esponenziali, avremo una gamma di parametri (n, λ) . Se Z_1, Z_2, \dots, Z_n sono V.A. normali standard e indipendenti, allora la somma dei loro quadrati è una V.A. che prende il nome di **chi-quadro a n gradi di libertà**, e si indica con

$$X \sim \chi_n^2$$

Essa è riproducibile, nel senso che se X_1 e X_2 sono due chi-quadro indipendenti, $X_1 + X_2$ è una chi-quadro con $n_1 + n_2$ gradi di libertà. Per dimostrare questo fatto non è necessario ricorrere alle funzioni generatrici, perché dalla definizione è evidente che $X_1 + X_2$ è la somma dei quadrati di $n_1 + n_2$ normali standard indipendenti. Definiamo la quantità $\chi_{\alpha,n}^2$, con α reale tra 0 e 1:

$$P(X \geq \chi_{\alpha,n}^2) = \alpha$$

6.7.1 La relazione tra chi-quadro e gamma

Una chi-quadro ad n gradi di libertà, corrisponde a una gamma di parametri $(n/2, 1/2)$. La densità di probabilità è perciò data da:

$$f(x) = \frac{x^{n/2-1} e^{-x/2}}{2^{n/2} \Gamma(n/2)}$$

6.8 Le distribuzioni T

Se Z e C_n sono V.A. indipendenti, la prima normale std e la seconda chi-quadro con n gradi di libertà, allora la V.A. T_n , definita come

$$T_n := \frac{Z}{\sqrt{C_n/n}}$$

si dice avere distribuzione t con n gradi di libertà, denotata con

$$T_n \sim t_n$$

La densità delle distribuzioni t è simmetrica rispetto all'asse di ascissa 0. È possibile dimostrare che al crescere di n , la densità di t_n converge a quella della normale standard. Come sempre, calcoliamo valore atteso e varianza:

$$\begin{aligned} E[T_n] &= 0 \\ Var(T_n) &= \frac{n}{n-2} \end{aligned}$$

Si noti che, al crescere di n , la varianza decresce, convergendo a 1, cioè alla varianza della n.s. In analogia con quanto fatto per le chi-quadro, definiamo $t_{\alpha,n}$

$$P(T_n \geq t_{\alpha,n}) = \alpha$$

Dalla simmetria rispetto allo zero otteniamo che

$$-t_{\alpha,n} = t_{1-\alpha,n}$$

6.9 Le distribuzioni F

Se C_n e $C : m$ sono V.A. indipendenti, di tipo chi-quadro con n e m gradi di libertà, allora la V.A. $F_{n,m}$ definita da:

$$F_{n,m} := \frac{C_n/n}{C_m/m}$$

si dice avere distribuzione F con n e m gradi di libertà. Come sempre, definiamo

$$P(F_{n,m} > P_{\alpha,n,m}) = \alpha$$

Notiamo, con calcoli, che

$$\frac{1}{F_{\alpha,n,m}} = F_{1-\alpha,n,m}$$

6.10 Distribuzione logistica

Una V.A. continua si dice avere distribuzione **logistica** di parametri (μ, ν) , con $\nu > 0$, se la sua funzione di ripartizione è data da

$$F(x) = \frac{e^{(x-\mu)/\nu}}{1 + e^{(x-\mu)/\nu}} \quad \forall x \in \mathbb{R}$$

Bella eh? Cerchiamo il valore atteso, e scopriamo che $E[X] = \mu$. ν è invece detta *dispersione*. Una V.A. logistica con media 0 e dispersione 1 è detta logistica standard.

7 La distribuzione delle statistiche campionarie

Se dovessimo studiare le caratteristiche di una popolazione molto grande, risulta ovvia l'impossibilità di analizzarne ogni singolo individuo. Si procede quindi solitamente a prenderne un campione casuale, e farvi inferenza. Un insieme X_1, X_2, \dots, X_n di V.A. indipendenti, tutte con la stessa distribuzione F , si dice campione o campione aleatorio della distribuzione F .

7.1 La media campionaria

Consideriamo una popolazione di elementi, a ognuno dei quali è associata una grandezza numerica. Denotiamo con μ e σ^2 la media e la varianza. Definiamo la media campionaria come

$$\bar{X} := \frac{X_1 + X_2 + \dots + X_n}{n}$$

Si noti che \bar{X} è una funzione delle variabili aleatorie $X_1 \dots X_n$. In quanto tale è una statistica, ed è una V.A. Ha senso quindi chiedersi quanto valgano il suo valore atteso e la sua varianza.

$$\begin{aligned} E[\bar{X}] &= \mu \\ \text{Var}(\bar{X}) &= \frac{\sigma^2}{n} \end{aligned}$$

Ha quindi lo stesso valore atteso della distribuzione da stimare, mentre la sua varianza è ridotta di un fattore n .

7.2 Il teorema del limite centrale

Affrontiamo ora uno dei risultati più notevoli della teoria della probabilità, ossia il **teorema del limite centrale**. In termini semplicistici, esso afferma che la somma di un numero elevato di V.A. indipendenti tende ad avere distribuzione approssimativamente normale. Di seguente l'enunciato: Siano X_1, X_2, \dots, X_n delle variabili aleatorie i.i.d., con media μ e varianza σ^2 . Allora, se n è grande, la somma $X_1 + X_2 + \dots + X_n$ è approssimativamente normale con media $n\mu$ e varianza $n\sigma^2$. Si può anche normalizzare la somma precedente allo scopo di ottenere una normale standard:

$$\frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \sim \mathcal{N}(0, 1)$$

in realtà sopra al \sim andrebbe un pallino, ad indicare "è approssimativamente distribuito come".

7.2.1 Distribuzione approssimata della media campionaria

Sia X_1, X_2, \dots, X_n un campione proveniente da una popolazione di media μ e varianza σ^2 . Il teorema appena enunciato ci permette di approssimare la distribuzione della media campionaria. Siccome il prodotto di una V.A. normale per una costante è ancora normale, ne segue che, quando n è grande, \bar{X} è approssimativamente gaussiana. Poiché, inoltre la media campionaria ha valore atteso μ e deviazione standard σ/\sqrt{n} , otteniamo che

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

7.3 La varianza campionaria

Sia X_1, X_2, \dots, X_n un campione aleatorio, proveniente da una distribuzione di media μ e varianza σ^2 . Sia \bar{X} la sua media campionaria. Introduciamo la varianza campionaria come

$$S^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

La sua radice quadrata S prende invece il nome di deviazione standard campionaria. Volendo calcolare $E[S^2]$ otteniamo che coincide con la varianza della popolazione.

$$E[S^2] = \sigma^2$$

Cerchiamo ora il valore atteso. Otteniamo che, come ci aspetteremmo, il valore atteso

$$E[S^2] = \sigma^2$$

7.4 Le distribuzioni delle statistiche di popolazioni normali

Osserviamo i casi in cui la popolazione è di tipo normale. Sia X_1, \dots, X_n un campione estratto da una distribuzione normale, di media μ e varianza σ^2 . Le V.A. sono indipendenti e $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Denotiamo come sempre

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Vogliamo trovare le loro distribuzioni.

7.4.1 La distribuzione della media campionaria

Siccome la somma di normali indipendenti è normale, \bar{X} la è. La sua media e la sua varianza, come nel caso generale, sono μ e σ^2/n e quindi

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

è una normale standard.

7.4.2 La distribuzione congiunta di \bar{X} e S^2

Vogliamo ora derivare la distribuzione di S^2 . Dimostriamo tramite l'identità che afferma che dati dei numeri x_1, \dots, x_n , posto $y_i := x_i - \bar{y}$

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n\bar{y}^2$$

trasformandola in:

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

che la distribuzione seguente è una chi-quadro:

$$(n-1) \frac{S^2}{\sigma^2} \sim \chi_{n-1}^2$$

Sappiamo inoltre che dato un campione X_1, \dots, X_n estratto da una popolazione gaussiana di media μ , se \bar{X} e S^2 denotano media e varianza campionaria, allora

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

Quindi, se si normalizza \bar{X} sottraendo la sua media μ e dividendo per la sua deviazione standard σ/\sqrt{n} , si ottiene una normale standard. Se invece dividiamo per S/\sqrt{n} si ha una distribuzione t con $n-1$ gradi di libertà.

7.5 Campionamento da insiemi finiti

Consideriamo una popolazione da N elementi. Con il concetto di **campione aleatorio** si intende la scelta di un sottoinsieme fatto in modo tale che tutti i sottoinsiemi $\binom{N}{n}$ abbiano la stessa possibilità di essere estratti. Supponiamo ora che una frazione p della popolazione abbia una determinata caratteristica. Avremo quindi pN elementi con la caratteristica, e $(1-p)N$ senza. Definiamo una variabile aleatoria

$$X_i := \begin{cases} 1 & \text{se l'elemento } i \text{ possiede la caratteristica} \\ 0 & \text{se l'elemento non possiede la caratteristica} \end{cases}$$

Consideriamo la somma di queste V.A.

$$X := X_1 + X_2 + \dots + X_n$$

Passiamo ora ad analizzare le probabilità associate alle statistiche X e \bar{X} . Notiamo innanzitutto che $P(X_i = 1) = p$. Notiamo che le variabili X_i sono bernoulliane **non indipendenti**. Ciò è ovvio perché, avendo estratto un elemento positivo, le probabilità della seguente estrazione diminuiscono:

$$P(X_2 = 1 | X_1 = 1) = \frac{pN-1}{N-1}$$

ci sono infatti $pN-1$ elementi positivi rimasti su $N-1$. È facile notare che se n è grande, la probabilità cambia poco e quindi si tende all'indipendenza. Siccome la somma di bernoulliane indipendenti è una binomiale, otteniamo che con n elevato, $X := \sum_i X_i$ è una binomiale di parametri n e p .

8 Stima parametrica

Consideriamo un campione aleatorio X_1, \dots, X_n estratto da una distribuzione F_0 che dipende da un vettore di parametri incogniti θ . Potrebbe ad esempio trattarsi di V.A. di Poisson, delle quali ignoriamo il valore di λ . Ci occupiamo ora di trovare dei metodi per stimare questi parametri incogniti, sia in maniera puntuale, che tramite i cosiddetti *intervalli di confidenza*.

8.1 Stimatori di massima verosimiglianza (ML)

Una qualunque statistica che cerchi di dare una stima di un parametro θ si dice **stimatore di θ** . Ad esempio, la media campionaria costituisce lo stimatore abituale della media μ di una distribuzione. Consideriamo delle variabili aleatorie X_1, \dots, X_n la cui distribuzione congiunta sia nota a meno di un parametro incognito θ . Per esempio, queste variabili hanno media θ incognita. Il nostro compito è quello di stimare θ partendo dai valori osservati. Utilizzeremo quindi uno *stimatore di massima verosimiglianza*, ottenibile con il ragionamento seguente. Denotiamo con $f(x_1, \dots, x_n | \theta)$ la funzione di massa congiunta di X_1, \dots, X_n , mostrando esplicitamente che f dipende da θ . Se interpretiamo f come la verosimiglianza che si realizzi la n-upla di dati x_1, \dots, x_n quando θ è il vero valore assunto dal parametro, sembra ragionevole adottare come stima il valore che rende massima la funzione. Conviene spesso usare il fatto che il logaritmo della suddetta ha il massimo nello stesso punto.

8.1.1 Stimare la distribuzione dei tempi di vita

Prendiamo una popolazione i cui elementi hanno un tempo di vita casuale. Sia X il numero di anni che vivrà una persona a caso nata oggi. Indichiamo con λ_i la probabilità che una persona nata oggi, e sopravvissuta per i primi $i - 1$ anni di vita, muoia durante l' i -esimo anno di vita.

$$\lambda_i := P(X = i | X > i - 1) = \frac{P(X = i)}{P(X > i - 1)}$$

poi, con $s_i = 1 - \lambda_i$, indichiamo la probabilità che sopravviva anche all'anno i . Le quantità λ_i e s_i sono dette rispettivamente funzione di rischio e funzione di sopravvivenza di un individuo che entra nel suo i -esimo anno di vita. Notiamo che

$$P(X = n) = P(X > n - 1) \lambda_n = s_1 s_2 \dots s_{n-1} \lambda_n$$

È perciò possibile stimare la funzione di massa di X a partire dalle stime delle quantità s_1, \dots, s_n , ottenibili da un campione di individui, considerando quelli che sono entrati nel loro i -esimo anno di vita un anno fa, e denotando con \hat{s}_i la frazione di essi ancora in vita oggi.

8.2 Intervalli di confidenza

Sia X_1, \dots, X_n un campione estratto da una popolazione normale di media incognita μ e varianza nota σ^2 . Abbiamo in precedenza dimostrato che $\bar{X} := \sum_i X_i / n$ è lo stimatore ML per μ . A volte, però, vorremmo un intervallo, piuttosto che un singolo punto. Per ottenerlo, dobbiamo fare uso della distribuzione di probabilità dello stimatore. Ricordando che \bar{X} è una normale, sappiamo che

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim \mathcal{N}(0, 1)$$

Perciò,

$$P\left(-1.96 < \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} < 1.96\right) \rightarrow P\left(\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}\right)$$

il 95% delle volte μ starà a una distanza non superiore a $1.96\sigma/\sqrt{n}$. Stiamo quindi affermando che nel 95% dei casi, la media apparterrà all'intervallo $\left(-1.96 \frac{\sigma}{\sqrt{n}}, +1.96 \frac{\sigma}{\sqrt{n}}\right)$. A volte ci interessa trovare un intervallo *unilaterale*. Per trovarlo, notiamo che se Z è $\mathcal{N}(0, 1)$, allora

$$0.95 \approx P(Z < 1.645) \approx P\left(\bar{X} - 1.645 \frac{\sigma}{\sqrt{n}} < \mu\right)$$

Avremo quindi un intervallo unilaterale destro con il 95% di confidenza

$$\left(\bar{x} - 1.645 \frac{\sigma}{\sqrt{n}}, \infty \right)$$

Possiamo ottenere livelli di confidenza per ogni livello. Ricordiamo infatti i numeri z_α tali che $P(Z > z_\alpha) = \alpha$. Questo implica che per ogni $\alpha \in (0, 1)$

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

Da cui deduciamo la formula generale

$$P\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Uguualmente, per gli intervalli unilaterali ci basterà usare z_α e non $z_{\alpha/2}$.

8.2.1 Intervalli di confidenza per la media di una normale, con varianza incognita

Abbiamo un campione X_1, \dots, X_n di una popolazione $\mathcal{N}(\mu, \sigma^2)$ con entrambi i parametri incogniti. Siccome la deviazione standard non è più nota, non possiamo più basarci sul fatto che $\sqrt{n}(\bar{X} - \mu)/\sigma$ è una normale standard. Tuttavia, usando la varianza campionaria S^2 , otteniamo una V.A. di tipo t con $n - 1$ gradi di libertà. Allora, poiché la densità delle distribuzioni t è simmetrica rispetto a 0 come quella della normale standard, avremo per $\alpha \in (0, 1/2)$

$$P\left(\bar{X} - t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}}\right)$$

8.2.2 Intervalli di confidenza per la varianza di una distribuzione normale

Abbiamo il solito campione proveniente da una normale, con entrambi i parametri incogniti. Possiamo costruire intervalli di confidenza per σ^2 basandoci sul fatto che

$$(n-1) \frac{S^2}{\sigma^2} \sim \chi_{n-1}^2$$

Infatti:

$$P\left(\frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}, n-1}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}, n-1}^2}\right)$$

8.3 Stime per la differenza tra le medie di due normali

Abbiamo due campioni X_1, \dots, X_n e Y_1, \dots, Y_n con parametri (μ_1, σ_1^2) e (μ_2, σ_2^2) . Vogliamo stimare $(\mu_1 - \mu_2)$. Dal momento che i due stimatori delle medie \bar{X} e \bar{Y} sono normali, assumiamo che $\bar{X} - \bar{Y}$ sia una normale. Inoltre,

$$\bar{X} - \bar{Y} \sim \left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} \right)$$

che abbiamo trovato sfruttando le proprietà di valore atteso e varianza. Ipotizzando di conoscere σ_1^2 e σ_2^2 , abbiamo che

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}} \sim \mathcal{N}(0, 1)$$

Usando i soliti passaggi, otteniamo

$$P\left(\bar{X} - \bar{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} < \mu_1 - \mu_2 < \bar{X} - \bar{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right) = 1 - \alpha$$

Dai, sono sicuro che all'esame te la ricorderai. Senza'altro.

8.4 Intervalli di confidenza approssimati per la media di una Bernoulliana

Consideriamo una popolazione di oggetti, ognuno dei quali indipendentemente da tutti gli altri soddisfa certi requisiti con probabilità incognita p . Nel caso vengano testati n di questi oggetti, rilevando quanti di essi raggiungono tali requisiti, come posso arrivare a un intervallo di confidenza per p ? Se X denota quanti oggetti, sugli n testati, soddisfano i requisiti, è intuitivo che essa sia una binomiale di parametri n e p . Quindi se n è sufficientemente elevato, essa si approssima a una normale.

$$\frac{X - np}{\sqrt{np(1-p)}} \sim \mathcal{N}(0, 1)$$

Quindi, come sempre

$$P\left(-z_{\frac{\alpha}{2}} < \frac{x - np}{\sqrt{np(1-p)}} < z_{\frac{\alpha}{2}}\right)$$

Tale regione non è però un intervallo. Se vogliamo ottenere un intervallo di confidenza vero e proprio, denotiamo con $\hat{p} := X/n$ la frazione di elementi che soddisfa i requisiti. Sappiamo che \hat{p} è lo stimatore di massima verosimiglianza di p , e ne è una buona approssimazione. Per questo,

$$\sqrt{n\hat{p}(1-\hat{p})} \approx \sqrt{np(1-p)}$$

Quindi, ne deduciamo che

$$\frac{X - np}{\sqrt{n\hat{p}(1-\hat{p})}} \sim \mathcal{N}(0, 1)$$

Che ci permette di ottenere un intervallo di confidenza

$$P\left(-z_{\frac{\alpha}{2}} < \frac{X - np}{\sqrt{n\hat{p}(1-\hat{p})}} < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

8.5 Intervalli di confidenza per la media di una distribuzione esponenziale

Consideriamo un campione X_1, \dots, X_n di variabili aleatorie esponenziali i.i.d. tutte con media θ incognita. Lo stimatore ML per la media è il solito. Per ottenere gli intervalli di confidenza, ricordiamo che $\sum_{i=1}^n X_i$ ha distribuzione gamma con parametri n e $1/\theta$. Sappiamo che

$$\frac{2}{\theta} \sum_{i=1}^n X_i \sim \chi_{2n}^2$$

Quindi per ogni $\alpha \in (0, 1)$

$$P\left(\frac{2 \sum_{i=1}^n X_i}{\chi_{\frac{\alpha}{2}, 2n}} < \theta < \frac{2 \sum_{i=1}^n X_i}{\chi_{1-\frac{\alpha}{2}, 2n}}\right)$$

8.6 Valutare l'efficienza degli stimatori puntuali

Sia $X := (X_1, \dots, X_n)$ un campione casuale estratto da una popolazione di distribuzione nota eccetto che per un parametro incognito θ , e sia $d = d(X)$ uno stimatore di θ . Come possiamo valutare l'efficacia di questo stimatore? Il criterio che utilizziamo è il cosiddetto *errore quadratico medio* dello stimatore, che per definizione è

$$r(d, \theta) := E[(d(X) - \theta)^2]$$

Sarebbe ideale avere un singolo stimatore d che minimizzasse $r(d, \theta)$ per tutti i valori di θ , ma ciò non è possibile. Cerchiamo quindi, ad esempio, stimatori che minimizzino $r(d, \theta)$ e siano non distorti. Definiamo

$$b_\theta(d) := E[d(X)] - \theta$$

il **bias** di d come stimatore di θ . Se esso è nullo, diremo che d è uno stimatore corretto, o *non distorto*. In altri termini, uno stimatore è corretto se il suo valore atteso coincide con il parametro che esso deve stimare.

8.7 Stimatori Bayesiani

Vista l'indeterminazione del parametro incognito θ , in alcune situazioni può essere ragionevole considerarlo assumere la forma di una V.A., facendo diventare il valore vero del parametro da stimare il numero realizzato dalla variabile. Questo approccio viene detto **bayesiano**, ed è di norma giustificato quando, prima di osservare gli esiti del campione di dati X_1, \dots, X_n , abbiamo delle informazioni sui possibili valori di θ e la loro plausibilità. Se queste informazioni assumono la forma di una distribuzione di probabilità, questa prende appropriatamente il nome di distribuzione a priori per θ . Supponiamo allora di poter esprimere le nostre considerazioni a priori su θ nella forma di una distribuzione continua, con densità di probabilità $p(\theta)$; osserviamo i valori di un campione di dati la cui distribuzione dipende da θ , e denotiamo con $f(x|\theta)$ la funzione di likelihood. Si tratta quindi della funzione di massa di probabilità/densità di probabilità, che esprime la plausibilità che uno dei dati sia uguale a x quando θ è il valore del parametro. Se i valori osservati sono $X_i = x_i$ allora la probabilità condizionale di θ è data da

$$f(\theta|x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n, \theta)}{f(x_1, \dots, x_n)}$$

La densità condizionale $f(\theta|x_1, \dots, x_n)$ è detta densità di probabilità a posteriori. Se conosciamo la distribuzione, la migliore stima di una V.A. è la sua media. Quindi, per stimare θ calcoliamo la media della distribuzione a posteriori $f(\theta|x_1, \dots, x_n)$. Tale stimatore è detto **stimatore bayesiano**, si indica con $E[\theta|X_1, \dots, X_n]$ e il suo valore si calcola nel modo usuale:

$$E[\theta|X_1, \dots, X_n] = \int_{-\infty}^{\infty} \theta f(\theta|x_1, \dots, x_n) d\theta$$

9 Verifica delle ipotesi

Abbiamo ora un campione aleatorio che ci è noto tranne che per uno o più parametri incogniti. Vogliamo verificare qualche ipotesi che li coinvolga. A priori, non sappiamo se l'ipotesi è vera o meno; vogliamo solo verificare se i dati raccolti la escludono, o no.

9.1 Livelli di significatività

Consideriamo una popolazione avente distribuzione F_θ che dipende da un parametro incognito θ , e supponiamo di voler verificare una qualche ipotesi su θ , che chiameremo *ipotesi nulla* e denoteremo con H_0 . Parleremo di *ipotesi semplice*, quando essa caratterizza completamente la distribuzione, di *ipotesi composta* quando non avviene. Supponiamo di voler disporre di un campione aleatorio X_1, \dots, X_n proveniente da questa popolazione, e di volerlo utilizzare per eseguire una verifica su una certa ipotesi nulla H_0 . Siccome dobbiamo valutare esclusivamente sugli n valori dei dati, il test sarà definito da una regione C nello spazio a n dimensioni, con l'intesa che se il vettore (X_1, \dots, X_n) cade all'interno di C l'ipotesi viene rifiutata, mentre viene accettata altrimenti. Una regione C con queste caratteristiche viene detta *regione critica* del test. Per esempio, se volessimo verificare che una popolazione gaussiana di varianza 1 abbia media 1, produrremmo la regione critica seguente:

$$C = \left\{ (X_1, \dots, X_n) : \left| 1 - \frac{1}{n} \sum_{i=1}^n X_i \right| > \frac{1.96}{\sqrt{n}} \right\}$$

Possono esistere due tipi di errori in questo procedimento:

- *Errori di prima specie*: quelli che gli informatici (*gente a posto*) chiamano **false negative**.
- *Errori di seconda specie*: quelli che gli informatici chiamano **false positive**.

Per bilanciare tra i due, specifichiamo un livello α , detto **livello di significatività**, e imponendo che il test abbia la proprietà che quando l'ipotesi H_0 è vera, la probabilità che venga rifiutata non possa superare α . Esso viene fissato solitamente attorno a 0.1, 0.05, 0.005.

9.2 La verifica delle ipotesi sulla media di una popolazione normale

9.2.1 Il caso in cui la varianza è nota

Supponiamo il solito campione aleatorio con parametri μ e σ^2 , con varianza nota e media incognita. Vogliamo verificare l'ipotesi nulla

$$H_0 : \mu = \mu_0$$

contro l'ipotesi alternativa

$$H_1 : \mu \neq \mu_0$$

Siccome lo stimatore puntuale normale per μ è $\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i$, è chiaro che dovremo accettare H_0 quando μ_0 non è troppo lontano da \bar{X} . La regione critica sarà quindi

$$C = \{ (X_1, \dots, X_n) : |\bar{X} - \mu_0| > c \}$$

scegliendo opportunamente la costante c . Se vogliamo un livello di significatività α , dobbiamo individuare quel valore di c che rende pari ad α la probabilità di errori di prima specie. Ciò significa che c deve soddisfare la relazione seguente:

$$\alpha = P_{\mu_0}(|\bar{X} - \mu_0| > c)$$

dove P_{μ_0} sta ad intendere che la probabilità viene calcolata con l'assunzione che $\mu = \mu_0$. Infatti la definizione di errore di prima specie significa che l'errore ci porta a rifiutare un'ipotesi in realtà vera. Quando però $\mu = \mu_0$, sappiamo che \bar{X} ha distribuzione normale con media μ e varianza σ^2/n , e quindi se Z denota una V.A. normale std, allora

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim Z$$

Otteniamo allora

$$P\left(|Z| > \frac{c\sqrt{n}}{\sigma}\right) = 2P\left(Z > \frac{c\sqrt{n}}{\sigma}\right)$$

Ed in definitiva, con il test con livello di significatività α

$$\begin{aligned} \text{si rifiuta } H_0 & \text{ se } \left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| > z_{\frac{\alpha}{2}} \\ \text{si accetta } H_0 & \text{ se } \left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| \leq z_{\frac{\alpha}{2}} \end{aligned}$$

9.2.2 I test unilaterali

Nel verificare l'ipotesi nulla $\mu = \mu_0$, abbiamo costruito un test che porta ad un rifiuto quando \bar{X} è lontana da μ_0 , ovvero, valori di \bar{X} troppo bassi o troppo elevati rispetto a μ_0 sembrano smentire che μ sia proprio uguale a μ_0 . Cosa accade invece quando μ può essere solo maggiore a μ_0 , quando non sono uguali? Ovvero, cosa occorre fare se l'ipotesi alternativa a $H_0 : \mu = \mu_0$ è $H_1 : \mu > \mu_0$? Dovremmo quindi rifiutare l'ipotesi nulla quando \bar{X} , lo stimatore di μ , è molto maggiore di μ_0 , e quindi la regione critica dovrebbe essere del tipo seguente:

$$C := \{(X_1, \dots, X_n) : \bar{X} - \mu_0 > c\}$$

per una scelta opportuna della costante c . In particolare, siccome la probabilità di rifiuto dovrebbe essere α quando H_0 è vera, occorre che c soddisfi la relazione

$$P_{\mu_0}(\bar{X} - \mu_0 > c) = \alpha$$

Di nuovo, poiché supponiamo $\mu = \mu_0$, esiste una statistica Z normale standard così definita:

$$Z := \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

Otteniamo quindi

$$P\left(Z > \frac{c\sqrt{n}}{\sigma}\right)$$

E, finalmente, ricaviamo c :

$$c = z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

Quella trovata è una regione critica **unilaterale**.

9.2.3 Quando la varianza non è nota: il test t

Fino ad ora abbiamo supposto di conoscere la varianza. Spesso non è così. Possiamo però sfruttare i calcoli precedenti, utilizzano, al posto della deviazione standard, il suo stimatore: la deviazione standard campionaria.

$$S := \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

rifiutando l'ipotesi nulla quando

$$\left| \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \right|$$

è troppo grande. Per definire questa "grandezza" sfruttiamo il fatto che questa assume una distribuzione t con $n - 1$ gradi di libertà. Denotiamola con T , e imponiamo che la probabilità di errori di prima specie sia α :

$$P_{\mu_0} \left(-c \leq \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \leq c \right)$$

Per ricavare c , sappiamo che la distribuzione t è simmetrica rispetto a 0, per cui $P(T > c) = \frac{\alpha}{2}$ e quindi $C = t_{\frac{\alpha}{2}, n-1}$. In definitiva:

$$\begin{aligned} \text{si rifiuta } H_0 & \text{ se } \left| \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \right| > t_{\frac{\alpha}{2}, n-1} \\ \text{si accetta } H_0 & \text{ se } \left| \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \right| \leq t_{\frac{\alpha}{2}, n-1} \end{aligned}$$

9.3 Verifica se due popolazioni normali hanno la stessa media

Vogliamo verificare se, ad esempio, due diversi approcci ad un problema hanno ricevuto risposte simili. Come fare? Possiamo verificare se "hanno la stessa media".

9.3.1 Se le varianze sono note

Abbiamo due campioni X_1, \dots, X_n e Y_1, \dots, Y_n indipendenti, provenienti da popolazioni normali di medie incognite μ_x e μ_y , e varianze note σ_x^2, σ_y^2 . Vogliamo verificare l'ipotesi

$$H_0 : \mu_x = \mu_y \text{ contro } H_1 : \mu_x \neq \mu_y$$

Sfruttando i due stimatori delle medie, sembra ragionevole rifiutare l'ipotesi quando $\bar{X} - \bar{Y}$ è molto lontano da 0. Troviamo un valore di c adatto, ricordando che

$$\bar{X} - \bar{Y} \sim \mathcal{N} \left(\mu_x - \mu_y, \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m} \right)$$

Per cui, per avere una normale standard:

$$\frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}} \sim \mathcal{N}(0, 1)$$

Quindi, quando H_0 è vera, la statistica del test $\frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}}$ ha distribuzione normale standard. Per cui, come sempre

$$\begin{aligned} \text{si rifiuta } H_0 & \text{ se } \frac{|\bar{X} - \bar{Y}|}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}} > z_{\frac{\alpha}{2}} \\ \text{si accetta } H_0 & \text{ se } \frac{|\bar{X} - \bar{Y}|}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}} \leq z_{\frac{\alpha}{2}} \end{aligned}$$

9.3.2 Se le varianze non sono note, ma si suppongono uguali

Con il caso di cui sopra, non sappiamo ora le varianze, ma sappiamo che sono uguali. Calcoliamo le due varianze campionarie, e ricordiamo che

$$\frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{S_p \sqrt{1/n + 1/m}} \sim t_{n+m-2}$$

dove S_p^2 è lo stimatore *pooled* di σ^2 , definito da

$$S_p^2 := \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}$$

Perciò, quando H_0 è vera, la statistica del test

$$T := \frac{\bar{X} - \bar{Y}}{S_p \sqrt{1/n + 1/m}} \sim t_{n+m-2}$$

ha distribuzione t con $n+m-2$ gradi di libertà. Ne segue che possiamo verificare l'ipotesi come

si rifiuta H_0 se $|T| > t_{\frac{\alpha}{2}, n+m-2}$

si accetta H_0 se $|T| \leq t_{\frac{\alpha}{2}, n+m-2}$

Potremmo anche determinare il p-dei-dati = $2P(T_{n+m-2} \geq |v|)$, dove T è una t di Student con $n+m-2$ gradi di libertà. Se invece volessimo verificare un'ipotesi unilaterale come

$$H_0 : \mu_x \leq \mu_y \text{ contro } H_1 : \mu_x > \mu_y$$

allora H_0 deve essere rifiutata per valori elevati di T , e in particolare il test ha livello di significatività α quando

si rifiuta H_0 se $T > t_{\alpha, n+m-2}$ si accetta H_0 se $T \leq t_{\alpha, n+m-2}$

Se v è il valore assunto dalla statistica T , allora il p-dei-dati corrispondente è $P(T_{n+m-2} \geq v)$.

9.3.3 Se le varianze sono ignote e diverse

In questo caso, usiamo gli stimatori al posto delle deviazioni standard. C'è una sola condizione a ciò: n e m devono essere molto elevati, in modo che la statistica

$$\frac{\bar{X} - \bar{Y}}{\sqrt{S_x^2/n + S_y^2/m}}$$

sia approssimativamente normale. Riassumendo:

$$\text{si accetta } H_0 \text{ se } -z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \bar{Y}}{\sqrt{S_x^2/n + S_y^2/m}} \leq z_{\frac{\alpha}{2}}$$

si accetta H_0 in tutti gli altri casi

9.4 La verifica di ipotesi sulla varianza di una popolazione normale

Abbiamo il solito campione, con media μ e varianza σ^2 , entrambe incognite. Vogliamo verificare l'ipotesi

$$H_0 : \sigma^2 = \sigma_0^2 \text{ contro } H_1 : \sigma^2 \neq \sigma_0^2$$

per un valore di σ_0^2 fissato. Per ottenere un test, ricordiamo che $(n-1)\frac{S^2}{\sigma_0^2}$ ha distribuzione chi-quadro con $n-1$ gradi di libertà. Perciò,

$$P_{H_0} \left(\chi_{1-\frac{\alpha}{2}, n-1}^2 \leq (n-1) \frac{S^2}{\sigma_0^2} \leq \chi_{\frac{\alpha}{2}, n-1}^2 \right) = 1 - \alpha$$

e in definitiva

$$\begin{aligned} \text{si accetta } H_0 & \text{ se } \chi_{1-\frac{\alpha}{2}, n-1}^2 \leq (n-1) \frac{S^2}{\sigma_0^2} \leq \chi_{\frac{\alpha}{2}, n-1}^2 \\ & \text{si rifiuta negli altri casi} \end{aligned}$$

Il p-dei-dati è il seguente

$$\text{p-dei-dati} = 2\min\{P(\chi_{n-1}^2 \leq c), 1 - P(\chi_{n-1}^2 \leq c)\}$$

9.4.1 Verificare se due popolazioni normali hanno la stessa varianza

Abbiamo i soliti due campioni, provenienti da popolazioni di parametri incogniti μ_x, σ_x^2 e μ_y, σ_y^2 , e verifichiamo

$$H_0 : \sigma_x^2 = \sigma_y^2 \text{ contro } H_1 : \sigma_x^2 \neq \sigma_y^2$$

Definiamo le varianze campionarie come al solito, e sappiamo che $(n-1)S_x^2/\sigma_x^2$ è una chi-quadro con $n-1$ gradi di libertà. Quindi, $(S_x^2/\sigma_x^2)/(S_y^2/\sigma_y^2)$ ha distribuzione F con parametri $n-1$ e $m-1$, e quando H_0 è vera,

$$\frac{S_x^2}{S_y^2} \sim F_{n-1, m-1}$$

da cui si deduce che

$$P_{H_0} \left(F_{1-\frac{\alpha}{2}, n-1, m-1} \leq \frac{S_x^2}{S_y^2} \leq F_{\frac{\alpha}{2}, n-1, m-1} \right) = 1 - \alpha$$

Si accetta quindi H_0 se $F_{1-\frac{\alpha}{2}, n-1, m-1} \leq \frac{S_x^2}{S_y^2} \leq F_{\frac{\alpha}{2}, n-1, m-1}$. In alternativa, si calcola il valore v assunto dalla statistica del test, $\frac{S_x^2}{S_y^2}$. Il p-dei-dati è uguale a $2\min\{P(F_{n-1, m-1} \leq v), 1 - P(F_{n-1, m-1} \leq v)\}$.

9.5 La verifica di ipotesi su una popolazione di Bernoulli

La distribuzione binomiale compare frequentemente nei problemi dell'ingegneria. Un esempio tipico è un processo produttivo che restituisce parti difettose o accettabili. Sappiamo che i pezzi saranno accettabili con probabilità p . Consideriamo un'ipotesi

$$H_0 : p \leq p_0 \text{ contro } H_1 : p > p_0$$

dove p_0 è un valore assegnato. Se denotiamo con X il numero di pezzi difettosi di un campione di n , dobbiamo certamente rifiutare H_0 quando X è troppo grande. Per calcolare quanto grande, notiamo che quando H_0 è vera

$$P_{H_0}(X \geq k) \leq \sum_k^n \binom{n}{i} p_0^i (1-p_0)^{n-i}$$

Per verificare le ipotesi ad un livello α , si deve rifiutare quando $X \geq k^*$ con k il numero più piccolo intero k tale che $\sum_{i=k}^n \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq \alpha$. Un modo migliore per implementare il test sarebbe ricavare prima il valore x della statistica del test, e poi calcolare il p-dei-dati come segue:

$$\text{p-dei-dati} = \sum_{i=x}^n \binom{n}{i} p_0^i (1-p_0)^{n-i}$$

9.5.1 Verificare se due popolazioni di Bernoulli hanno lo stesso parametro

Immaginiamo di voler confrontare due diversi metodi di fabbricazione per transistor. Indichiamo con p_1 e p_2 le probabilità, incognite, che un pezzo prodotto con i metodi 1 e 2 sia difettoso. Raccogliamo poi i campioni di numerosità n_1 e n_2 e indichiamo con X_1 e X_2 i pezzi difettosi trovati. Otterremo un'ipergeometrica di parametri n_1, n_2, k , con $k = X_1 + X_2$. Perciò,

$$P_{H_0}(X_1 = i | X_1 + X_2 = k) = \frac{\binom{n_1}{i} \binom{n_2}{k-i}}{\binom{n_1+n_2}{k}}$$

Volendo realizzare il test

$$H_0 : p_1 = p_2 \text{ contro } H_1 : p_1 \neq p_2$$

si osserva quanto valgono $X_1 = x_1$ e $X_2 = x_2$ e si calcola la somma $k = x_1 + x_2$. Denotata quindi X una V.A. ipergeometrica, se

$$\begin{aligned} \text{si rifiuta } H_0 \text{ se } P(X \leq x_1) < \frac{\alpha}{2} \text{ o } P(X \geq x_1) < \frac{\alpha}{2} \\ \text{si accetta altrimenti} \end{aligned}$$

Il p-dei-dati è pari a $2\min\{P(X \leq x_1), P(X \geq x_1)\}$.

9.6 La verifica di ipotesi sulla media di una distribuzione di Poisson

Sia X una Poisson con media λ , e supponiamo di voler confrontare le ipotesi

$$H_0 : \lambda = \lambda_0 \text{ contro } H_1 : \lambda \neq \lambda_0$$

Se x è il valore osservato per X , un test con controllo di significatività α deve rifiutare l'ipotesi nulla se

$$P_{\lambda_0}(X \geq x) \leq \frac{\alpha}{2} \text{ o se } P_{\lambda_0}(X \leq x) \leq \frac{\alpha}{2}$$

dove con P_{λ_0} intendiamo la probabilità calcolata assumendo che X abbia media λ_0 . Il p-dei-dati è equivalente a $2\min\{P_{\lambda_0}(X \leq x), P_{\lambda_0}(X \geq x)\}$.

9.6.1 Testare la relazione tra i parametri di due Poisson

Siano X_1 e X_2 due V.A. di Poisson indipendenti, con medie rispettivamente λ_1 e λ_2 , e di voler vagliare l'ipotesi

$$H_0 : \lambda = c\lambda_1 \text{ contro } H_1 : \lambda_2 \neq c\lambda_1$$

per una costante c assegnata. Il test che costruiamo è di tipo condizionale, ed è basato sul fatto che la distribuzione di X_1 condizionata al valore della somma di $X_1 + X_2$, è binomiale. In particolare, la binomiale avrà parametri $(X_1 + X_2)$ e $\lambda_1/(\lambda_1 + \lambda_2)$. Dovremo allora rifiutare H_0 qualora

$$P(Y \geq x_1) \leq \frac{\alpha}{2} \text{ oppure } P(Y \leq x_1) \leq \frac{\alpha}{2}$$

dove Y è la binomiale.

10 Regressione

Spesso vogliamo determinare le relazioni tra variabili, piuttosto che informazioni sulle stesse. Sarà quindi necessario avere più ingressi x_1, \dots, x_r , ed una sola uscita Y . Il modello suppose che la risposta sia in funzione degli ingressi; per questo Y è anche detta variabile dipendente, mentre le x , indipendenti. La più semplice relazione che è possibile immaginare è quella lineare; essa si presenta quando per delle opportune costanti β_0, \dots, β_r vale l'equazione

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_r x_r$$

Se la relazione fosse questa saremo a posto. Visto che non la è, dovremo sommare alla suddetta un errore casuale e . Possiamo infatti esprimerla anche come

$$E[Y|\mathbf{x}] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_r x_r$$

Infine, enunciamo l'**equazione di regressione lineare**, che con i suoi **coefficienti di regressione** β si dice semplice se $r = 1$, multipla altrimenti.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_r x_r + e$$

O, in caso di modello semplice:

$$Y = \alpha + \beta x + e$$

10.1 Stima dei parametri di regressione

Supponiamo di osservare, per i che va da 1 a n , le risposte Y_i corrispondenti a certi valori di ingresso x_i e di volerle usare per stimare α e β in un modello di regressione lineare semplice. Se A e B sono gli stimatori cercati, allora $A + Bx_i$ è lo stimatore della risposta. Poiché la risposta realmente ottenuta è Y_i , la quantità $(Y_i - A - Bx_i)^2$ rappresenta il quadrato della differenza tra predizione e valore osservato, e dovrebbe quindi essere minimo. Denotiamo con SS la somma di questi dati:

$$SS := \sum_{i=1}^n (Y_i - A - Bx_i)^2$$

Il metodo dei minimi quadrati consiste nello scegliere come stimatori di α e β i due valori di A e B che minimizzano SS . Deriviamo SS rispetto ad A e B .

$$\begin{aligned} \frac{dSS}{dA} &= -2 \sum_{i=1}^n (Y_i - A - Bx_i) \\ \frac{dSS}{dB} &= -2 \sum_{i=1}^n x_i (Y_i - A - Bx_i) \end{aligned}$$

Per trovare il minimo, li poniamo a 0, ottenendo, dopo calcoli, gli stimatori dei minimi quadrati

$$B = \frac{\sum_i x_i Y_i - \bar{x} \sum_i Y_i}{\sum_i x_i^2 - n\bar{x}^2} \quad A = \bar{Y} - B\bar{x}$$

La retta $A + Bx_i$ è la stima della retta di regressione, ossia la retta che interpola meglio i dati.

10.2 Distribuzione dei dati

Se fino ad ora è stato sufficiente supporre che gli errori casuali avessero media nulla, non possiamo illuderci per sempre. Scopriamo, riscrivendo B , che esso è una combinazione lineare delle normali Y_1, \dots, Y_n , ed è quindi normale. Calcolandone i parametri, otteniamo che $E[B] = \beta$, e quindi B è uno stimatore non distorto. Passando ad A , otteniamo che $E[A] = \alpha$ ed anch'esso è uno stimatore corretto. Calcoliamo la varianza di A esprimendolo come combinazione lineare di Y_1, \dots, Y_n . Il risultato è che

$$Var(A) = \frac{\sigma^2 \sum_i x_i^2}{n (\sum_i x_i^2 - n\bar{x}^2)}$$

Ritornando ai nostri residui SS_R , osserviamo che

$$\frac{SS_R}{\sigma^2} \sim \chi_{n-2}^2$$

il fatto che sia una chi-quadro, implica che

$$E \left[\frac{SS_R}{\sigma^2} \right] = n - 2 \quad E \left[\frac{SS_R}{n - 2} \right] = \sigma^2$$

Così che $\frac{SS_R}{n-2}$ è uno stimatore non distorto di σ^2 . Definiamo ora una notazione che sarà comoda in futuro:

$$\begin{aligned} S_{xY} &:= \sum_1^n (x_i - \bar{x})(Y_i - \bar{Y}) \\ S_{xx} &:= \sum_1^n (x_i - \bar{x})^2 \\ S_{YY} &:= \sum_1^n (Y_i - \bar{Y})^2 \end{aligned}$$

Possiamo così riscrivere gli stimatori

$$B = \frac{S_{xY}}{S_{xx}} \quad A = \bar{Y} - B\bar{x} \quad SS_R = \frac{S_{xx}S_{YY} - S_{xY}^2}{S_{xx}}$$

10.3 Inferenza statistica sui parametri di regressione

Grazie al *refactoring* appena fatto, costruire test statistici e intervalli di confidenza è **relativamente** semplice.

10.3.1 Inferenza su β

Un'ipotesi importante da verificare è quella che β sia pari a zero. Di fatti, se è vera, la risposta non dipende dall'ingresso, ovvero non vi è correlazione tra le due variabili. Per verificare

$$H_0 : \beta = 0 \text{ contro } H_1 : \beta \neq 0$$

notiamo che

$$\frac{B - E[B]}{\sqrt{Var(B)}} = \frac{B - \beta}{\sigma / \sqrt{S_{xx}}} \sim \mathcal{N}(0, 1)$$

e inoltre tale V.A. è indipendente da

$$\frac{SS_R}{\sigma^2} \sim \chi_{n-2}^2$$

Perciò, dalla definizione di distribuzione t segue che

$$\frac{B - \beta}{\sigma/\sqrt{S_{xx}}} \sqrt{\frac{\sigma^2(n-2)}{SS_R}} = \sqrt{\frac{(n-2)S_{xx}}{SS_R}} (B - \beta) \sim t_{n-2}$$

Quindi, quando H_0 è vera,

$$\sqrt{\frac{(n-2)S_{xx}}{SS_R}} B \sim t_{n-2}$$

se volessimo avere un livello di significatività γ , quindi

$$\text{si rifiuta } H_0 \text{ se } \sqrt{\frac{(n-2)S_{xx}}{SS_R}} |B| > t_{\frac{\gamma}{2}, n-2}$$

si accetta H_0 negli altri casi

Si potrebbe anche calcolare il valore v assunto da $\sqrt{\frac{(n-2)S_{xx}}{SS_R}} |B|$ e rifiutando quindi H_0 quando il livello di significatività è maggiore o uguale a

$$\text{p-dei-dati} = P(|T_{n-2}| > v) = 2P(T_{n-2} > v)$$

10.3.2 Regressione alla media

Assumiamo che vi sia una relazione lineare tra il valore della stessa caratteristica in un figlio (Y) ed un genitore x . Si avrà una regressione verso la media ogni volta che il parametro β è compreso tra 0 e 1. Ovvero, se

$$E[Y] = \alpha + \beta x$$

e $0 < \beta < 1$, allora $E[Y]$ sarà più piccolo di x quando x è molto grande, e più grande di x quando x è molto piccolo.

10.3.3 Inferenza su α

La determinazione degli intervalli di confidenza e dei test statistici che riguardano il parametro α si ottiene in modo analogo a quanto fatto per β . In particolare, possiamo mostrare che

$$\sqrt{\frac{n(n-2)S_{xx}}{SS_R \sum_i x_i^2}} (A - \alpha) \sim t_{n-2}$$

di conseguenza, ad un livello $1 - \gamma$, l'intervallo di confidenza bilaterale è dato da

$$A \pm t_{\frac{\gamma}{2}, n-2} \sqrt{\frac{SS_R \sum_i x_i^2}{n(n-2)S_{xx}}}$$

10.3.4 Inferenza sulla risposta media $\alpha + \beta x$

Una questione certamente interessante è l'utilizzo delle coppie di dati $(x_i, Y_i), i = 1, 2, \dots, n$ per stimare $\alpha + \beta x_0$, vale a dire la risposta media per un livello di ingresso assegnato x_0 . Se si desidera uno stimatore puntuale, la scelta ovvia è $A + Bx_0$, che è uno stimatore non distorto, visto che A e B lo sono entrambi. Se invece vogliamo ottenere degli intervalli di confidenza, oppure verificare delle ipotesi sulla risposta media, è necessario prima determinare la distribuzione dello stimatore $A + Bx_0$, che scopriremo essere normale. La media già la conosciamo, ci serve la varianza

$$Var(A + Bx_0) = \sigma^2 \left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}} \right]$$

Quindi, in definitiva

$$A + Bx_0 \sim \mathcal{N} \left(\alpha + \beta x_0, \sigma^2 \left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}} \right] \right)$$

Non possiamo però fare inferenza su questa statistica perché σ^2 è incognita. Per questo, sfruttiamo il fatto che $A + Bx_0$ sia indipendente da

$$\frac{SS_R}{\sigma^2} \sim \chi_{n-2}^2$$

ricaviamo quindi gli intervalli di confidenza

$$A + Bx_0 \pm t_{\frac{\gamma}{2}, n-2} \sqrt{\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}}} \sqrt{\frac{SS_R}{n-2}}$$

10.3.5 Intervallo di predizione di una risposta futura

Ipotizziamo ora di voler trovare non un valore medio di $\alpha + \beta x$, quanto un singolo valore dato un ingresso specifico. Lo stimatore adatto sarebbe quello della sua media $A + Bx_0$. Se volessimo però un intervallo di valori che contenga la risposta con un determinato livello di confidenza? Denotiamo semplicemente Y, la risposta futura con un livello di ingresso x_0 , e consideriamo la distribuzione di probabilità di $Y - A - Bx_0$, cioè la differenza tra risposta e valore predetto. Sappiamo per ipotesi che Y è normale di parametri $\alpha + \beta x_0$ e σ^2 . Sappiamo inoltre da prima che

$$A + Bx_0 \sim \mathcal{N} \left(\alpha + \beta x_0, \sigma^2 \left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}} \right] \right)$$

Di conseguenza, la differenza è ancora una normale

$$Y - A - Bx_0 \sim \mathcal{N} \left(0, \sigma^2 \left[1 + \frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}} \right] \right)$$

Usando il fatto che SS_R è indipendente da A e B, come pure da Y, e che $\frac{SS_R}{\sigma^2}$ è una chi-quadro con $n - 2$ gradi di libertà, sostituiamo e otteniamo una t :

$$\frac{Y - A - Bx_0}{\sqrt{\frac{n+1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}}} \sqrt{\frac{SS_R}{n-2}}} \sim t_{n-2}$$

In definitiva, per ogni $0 < \gamma < 1$, si ha che

$$P \left(-t_{\frac{\gamma}{2}, n-2} < \frac{Y - A - Bx_0}{\sqrt{\frac{n+1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}}} \sqrt{\frac{SS_R}{n-2}}} < t_{\frac{\gamma}{2}, n-2} \right) = 1 - \gamma$$

10.4 Coefficiente di determinazione e coefficiente di correlazione campionaria

Supponiamo di voler esprimere la variabilità o dispersione dell'insieme di risposte Y_1, \dots, Y_n , ottenute con livelli di ingresso x_1, \dots, x_n . Una comune misura statistica della variabilità è costituita da

$$S_{YY} := \sum_1^n (Y_i - \bar{Y})^2$$

una quantità che rappresenta, a meno di un fattore moltiplicativo, la varianza campionaria delle Y_i . Se esse fossero ad esempio tutte uguali tra loro -e quindi tutte uguali a \bar{Y} - il valore di S_{YY} sarebbe pari a 0. Questa variabilità viene però da due contributi. In primis, sappiamo che se le x_i non sono tutte uguali, le Y_i hanno valori attesi diversi, e questo disperde le loro realizzazioni. In secundis, una volta che si tenga conto della variabilità delle x_i , ogni Y_i ha distribuzione con varianza σ^2 attorno al suo valore atteso e non coinciderà quindi esattamente con le nostre predizioni. Cerchiamo di quantificare quale parte della variabilità sia dovuta all'ingresso e quale alla varianza propria delle risposte. Notiamo che la quantità

$$SS_R := \sum_1^n (Y_i - A - Bx_i)^2$$

misura quella parte di variabilità intrinseca delle risposte. Di conseguenza, $S_{YY} - SS_R$ rappresenta l'altra parte, cioè quella che si spiega con la diversità dei livelli di ingresso. La statistica R^2 , detta *coefficiente di determinazione*, è la frazione della variabilità totale che si spiega con la diversità dei livelli di ingresso

$$R^2 := \frac{S_{YY} - SS_R}{S_{YY}} = 1 - \frac{SS_R}{S_{YY}}$$

10.5 Analisi dei residui: verifica del modello

Per verificare se un modello di regressione lineare semplice

$$Y = \alpha + \beta x + e$$

si adatti o meno ai dati, analizziamo i cosiddetti *residui*, $Y_i - (A + Bx_i)$. Per prima cosa essi vanno normalizzati, dividendoli per lo stimatore della deviazione standard delle Y_i :

$$\frac{Y_i - (A + Bx_i)}{\sqrt{SS_R/(n-2)}}$$

ottenendo i cosiddetti *residui standardizzati*. Quando il modello è corretto, questi residui sono approssimativamente normali standard indipendenti, quindi distribuiti attorno allo zero.

10.6 Linearizzazione

In certe situazioni può essere evidente che la risposta non sia una funzione lineare del livello di ingresso. Possiamo però a volte riportarci al caso lineare con un cambiamento di variabili. Ad esempio, immaginiamo di avere una relazione esponenziale. Potremmo linearizzarla prendendo i logaritmi, e poi stimare α e β con il metodo dei minimi quadrati.

10.7 Minimi quadrati pesati

Nel solito modello di regressione $Y = \alpha + \beta x + e$ può capitare che la varianza delle risposte non sia costante ma dipenda dal livello di ingresso. Se queste dipendenze sono note, oppure se sono note a meno di un fattore moltiplicativo, i parametri di regressione si possono stimare minimizzando una somma pesata dei residui al quadrato. In particolare, se

$$Var(Y_i) = \frac{\sigma^2}{w_i}$$

con le w_i note e σ^2 eventualmente ignota, allora gli stimatori A e B vanno scelti in modo da minimizzare

$$\sum_1^n \frac{[Y_i - (A + Bx_i)]^2}{Var(Y_i)} = \frac{1}{\sigma^2} \sum_1^n w_i (Y_i - A - Bx_i)^2$$

Calcolando le derivate parziali rispetto ad A e B, otteniamo il seguente sistema

$$\begin{cases} \sum_1^n w_i Y_i = A \sum_1^n w_i + B \sum_1^n w_i x_i \\ \sum_1^n w_i x_i Y_i = A \sum_1^n w_i x_i + B \sum_1^n w_i x_i^2 \end{cases}$$

Queste equazioni possono essere facilmente risolte per trovare gli stimatori dei minimi quadrati.

10.8 Regressione polinomiale

Nei casi in cui la relazione che lega la variabile di risposta Y con quella indipendente x non possa essere approssimata adeguatamente con modelli lineari, si può a volte prendere in considerazione le relazioni polinomiali, come

$$U = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_r x^r + e$$

dove i β sono i coefficienti di regressione che è necessario stimare. Supponendo che i dati consistano di n coppie di valori (x_i, Y_i) , gli stimatori dei minimi quadrati di β_0, \dots, β_r , che denotiamo con B_0, \dots, B_r , sono quei valori che rendono minima l'espressione seguente:

$$\sum_1^n (Y_i - B_0 - B_1 x_i - B_r x_i^r)^2$$

Per determinarli calcoliamo le derivate parziali rispetto a B_0, \dots, B_r e le poniamo uguali a 0. Otteniamo un sistema in $r + 1$ equazioni dette *equazioni normali*.

$$\begin{cases} \sum_1^n Y_i = B_0 n + B_1 \sum_1^n x_i + B_2 \sum_1^n x_i^2 + \dots + B_r \sum_1^n x_i^r \\ \sum_1^n x_i Y_i = B_0 \sum_1^n x_i + B_1 \sum_1^n x_i^2 + B_2 \sum_1^n x_i^3 + \dots + B_r \sum_1^n x_i^{r+1} \\ \dots \\ \sum_1^n x_i^r Y_i = B_0 \sum_1^n x_i^r + B_1 \sum_1^n x_i^{r+1} + B_2 \sum_1^n x_i^{r+2} + \dots + B_r \sum_1^n x_i^{2r} \end{cases}$$

La scelta del grado va svolta studiando il diagramma di dispersione, che spesso può darne un'idea.

10.9 Regressione lineare multipla

Nella gran parte delle applicazioni, la risposta può essere predetta in maniera migliore se, al posto di basarsi su una singola variabile, se ne utilizzano diverse. Studiamo il modello di regressione con k variabili indipendenti, la risposta è legata loro tramite una relazione lineare

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + e$$

dove per j che va da 1 a k , x_j è il livello della j -esima variabile di ingresso ed e è un errore casuale che assumeremo di media nulla e varianza costante incognita. I parametri β devono essere stimati dai dati, che consisteranno di n osservazioni di risposte Y_i unitamente ai rispettivi livelli di ingresso. Le variabili Y_i sono legate agli ingressi tramite

$$E[Y_i] = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$$

Se denotiamo con B_0, \dots, B_k gli stimatori di β , allora la somma dei residui al quadrato è

$$\sum_1^n (Y_i - B_0 - B_1 x_{i1} - \dots - B_k x_{ik})^2$$

ed è precisamente l'espressione da minimizzare dagli stimatori dei minimi quadrati. Per determinarli, facciamo le derivate parziali rispetto a B_i e le poniamo uguali a 0, giungendo al sistema seguente:

$$\begin{cases} \sum_1^n Y_i = B_0 n + B_1 \sum_1^n x_{i1} + B_2 \sum_1^n x_{i2} + \dots + B_k \sum_1^n x_{ik} \\ \sum_1^n x_{i1} Y_i = B_0 \sum_1^n x_{i1} + B_1 \sum_1^n x_{i1}^2 + B_2 \sum_1^n x_{i1} x_{i2} + \dots + B_k \sum_1^n x_{i1} x_{ik} \\ \dots \\ \sum_1^n x_{ik} Y_i = B_0 \sum_1^n x_{ik} + B_1 \sum_1^n x_{i1} x_{ik} + B_2 \sum_1^n x_{ik} x_{i2} + \dots + B_k \sum_1^n x_{ik}^2 \\ \dots \end{cases}$$

Conviene introdurre una notazione matriciale.

$$Y := \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{bmatrix} \quad X := \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \quad \beta := \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{bmatrix} \quad e := \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix}$$

Con questa notazione, il modello di regressione multipla può essere scritto nella forma

$$Y = X\beta + e$$

Se inoltre denotiamo con

$$B := \begin{bmatrix} B_0 \\ B_1 \\ \dots \\ B_k \end{bmatrix}$$

allora le equazioni normali prendono la forma

$$X'XB = X'Y$$

dove X' è la trasposta di X . Se poi $X'X$ è invertibile, possiamo ricavare gli stimatori dei minimi quadrati tramite

$$B = (X'X)^{-1} X'Y$$