

39 Test del χ^2 di indipendenza

Troviamo in un articolo⁽¹⁹⁴⁾ scientifico 2 test del chi quadrato, relativi alle stesse 134 persone ma stratificate con 2 diverse soglie per la carenza di vitamina D:

Our study demonstrates that only severe 25OHD deficiency at the beginning of COVID19 infection was associated with an increased subsequent risk of developing an overt infection

(semplificatamente: avere poca vitamina D non risulta molto associato alla malattia ma averne proprio pochissima sì) e dal medesimo articolo scientifico fotografiamo:

Table 2. Distribution analysis of subsequent cases of COVID-19 infections in healthy subjects during 12 months observation period, stratified according to vitamin D (25OHD) concentrations above or below 12 ng/mL (30 nmol/L) in spring 2020. The difference is statistically significant, $p = 0.0063$, χ^2 test.

COVID-19 Infection	25OHD < 12 ng/mL		
	NO	YES	
NO	89 (for > 12 ng/mL)	14 (for < 12 ng/mL)	103 (76.9%)
YES	20 (for > 12 ng/mL)	11 (for < 12 ng/mL)	31 (23.1%)
Total	109	25	134 (100%)

Table 3. Distribution analysis of subsequent cases of COVID-19 infections during 12 months observation period in healthy subjects, stratified according to vitamin D (25OHD) concentrations above or below 20 ng/mL (50 nmol/L) in spring 2020. The difference was not statistically significant, $p = 0.2787$, χ^2 test.

COVID-19 Infection	25OHD < 20 ng/mL		
	NO	YES	
NO	41 (for > 20 ng/mL)	62 (for < 20 ng/mL)	103 (76.9%)
YES	9 (for > 20 ng/mL)	22 (for < 20 ng/mL)	31 (23.1%)
Total	50	84	134 (100%)

39.1 Inquadramento della questione

In questa trattazione elementare, verrà considerato solo il caso più semplice, con 4 valori/dati. Fissiamo subito le idee su un esempio.

¹⁹⁴Basińska-Lewandowska M, Lewandowski K, Horzelski W, Lewiński A, Skowrońska-Józwiak E. Frequency of COVID-19 Infection as a Function of Vitamin D Levels. *Nutrients*. 2023 Mar 24;15(7):1581. doi: 10.3390/nu15071581. PMID: 37049423; PMCID: PMC10097275.

Supponiamo di avere 280 monete uguali di sconosciuta probabilità di dare testa, e che a 50 di esse venga fatta meccanicamente una piccola incisione al centro sulla faccia della testa. Le monete trattate hanno tutte un'uguale sconosciuta nuova probabilità di dare testa. (Di maggiore interesse potrebbe essere il trattamento con un farmaco, e l'eventuale morte a 5 anni, ma le monete e i dadi restano modelli generali ottimali, anche perché da essi *a priori ci aspettiamo* di default un comportamento casuale, mentre per un farmaco saremmo portati a supporre causalità – contenendo la tal molecola *dovrebbe* fare un certo effetto – eventualmente inesistenti: nella *statistica medica* dobbiamo appunto verificare se funziona, *contando i morti*, per così dire). Lanciamo una volta ciascuna delle 280 monete e contiamo le teste nelle 2 classi. Supponiamo di avere questa situazione (risultato=viene 1)

	non risultato	risultato
non trattati	188	42
trattati	43	7

Naturalmente $43 + 7$ sono le 50 monete trattate, e $188 + 42$ sono le rimanenti 230, non trattate. Il trattamento ha diminuito la probabilità di dare testa? Verrebbe da dire di sì, perché fra i trattati la frequenza relativa di 1 è $7/50 = 0.14$ mentre fra i non trattati è $42/230 = 0.183$ ma dobbiamo ragionevolmente escludere un effetto casuale: tale miglioramento potrebbe apparire per caso anche se a nostra insaputa il tecnico preposto non avesse affatto compiuto l'incisione dichiarata.

Abbiamo una variabile aleatoria X con 2 valori, "trattate" e "non trattate", e una variabile aleatoria Y con 2 valori, "testa" e "croce". Il test statistico è definito da

$$H : \text{indipendenza} \quad A : \text{non indipendenza}$$

(vorremmo escludere l'indipendenza, cioè che l'incisione non abbia prodotto risultato).

Come in tutti gli altri test statistici che consideriamo, si fissa un livello di confidenza α , tipicamente $0.95 = 95\%$, ovvero un livello di significatività $1 - \alpha$, tipicamente $0.05 = 5\%$.

(Si ritiene che per correttezza scientifica ovvero deontologica, questo valore vada fissato *prima* di fare i calcoli).

Si calcolerà uno *stimatore* (ovvero *statistica*) T , dalla formula alquanto complessa, e lo si confronterà col quantile $\chi^2_{1-\alpha}(k)$ del chi quadrato con

$$k := (m - 1) \cdot (n - 1) = (\#righe - 1) \times (\#colonne - 1) \quad (153)$$

gradi di libertà, (qua 1 perchè abbiamo 2 righe e 2 colonne) e se T supera il quantile

$$T > \chi^2_{1-\alpha}(k) \quad (154)$$

si respinge l'ipotesi, altrimenti non la si respinge (che alcuni dicono "si accetta"), al livello di significatività $1 - \alpha$ ovvero di confidenza α .

Con le tabelle 2×2 (*tetracoriche*) all'usuale livello del 95%, il quantile è ≈ 3.84 .

Il calcolo di T invece è più complesso e i vari Autori non concordano sulla formula, che per noi sarà la (156).

39.2 Semplice statistica per trial clinico

In pratica per la tabella 2×2 a 4 valori a, b, c, d , ultrasemplificando in *guariti* e *non guariti* i possibili esiti clinici, ecco il calcolo – che ci proponiamo di imparare a memoria – da fare per

	guariti	non guariti
farmaco 1	a	b
farmaco 2	c	d

al livello 0.05 ovvero 0.95 rifiuto l'ipotesi nulla

H : il farmaco 1 equivale al farmaco 2

se

(155)

"formula del chi quadrato" > 3.84

essendo la "formula del chi quadrato" (termine non standard)

$$T := \frac{(ad - bc)^2 (a + b + c + d)}{(a + b)(a + c)(b + d)(c + d)} \quad (156)$$

(che non ci proponiamo di imparare a memoria ma solo di saper usare se viene data).

(Altri Autori usano una formula leggermente diversa e più complessa con la *correzione di Yates*).

(Si tratta ovviamente dell'equivalenza dei farmaci dal solo punto di vista delle guarigioni, non del costo o di un'eventuale tossicità a lunga scadenza, e la guarigione andrà ben definita, in particolare nella sua determinazione temporale: lo stesso farmaco che elimina un sintomo in un giorno potrebbe uccidere in un mese, si ripensi all'eroina della Bayer).

Naturalmente il valore 3.84 è il quantile del chi quadrato (a 1 grado di libertà com'è dovuto per la tabella 2×2) al livello 0.05; e si potrà modificarlo per livelli diversi, in particolare 6.63 al livello 0.01, molto più probativo; si veda la tavola del chi quadrato in 48.4.

Usuale quantile del chi quadrato
d'indipendenza:
3.84

△ 8 □
3, 8 4

3 ← CHI

8 ← QUADRATO

4 ← TEST

Figure 57: Mnemonico per il numero 3.84: contare le lettere delle parole.

Esempio 1. Per un ipotetico antibiotico per cani, coi valori

	guariti	non guariti
xmicina	52	10
streptomicina	40	21

in http://www.quadernodiepidemiologia.it/epi/assoc/chi_qua.htm trovano $T = 5.46 > 3.84$ e concludono

consente di ritenere che la differenza fra i due gruppi sia significativa al livello di probabilità 5% ma non al livello di probabilità 1%.

E infatti immettendo i dati per esempio in <https://www.socscistatistics.com/tests/chisquare/default2.aspx>, dove c'è un calcolatore online del chi quadrato di indipendenza, troviamo che

chi-square statistic is 5.4607

e che il valore discriminante fra il rifiuto e il non rifiuto cioè il

p -value is .019449. Significant at $p < .05$. E si noti che 0.019 è sì $< 5\%$ ma non $< 1\%$.

Naturalmente il valore 5.46 si ottiene con la (156), anche con la calcolatrice.

L'intero calcolo può essere fatto online gratuitamente anche per tabelle più grandi in altri siti web, oltretutto con vari software statistici usati nelle Scienze Applicate.

Esempio 2. In http://www.biostatistica.unich.it/mat_didattica/Prof_di_nicola/3-test%20del%20chi%20quadrato.pdf invece che un farmaco e la possibile guarigione, vengono considerati un intossicante e una possibile malattia, e ovviamente la teoria è la stessa:

	bronchite	non bronchite
fumo	120	280
non fumo	93	527

$$T = \frac{(120 \cdot 527 - 93 \cdot 280)^2}{213 \cdot 807 \cdot 400 \cdot 620} \approx 32.21$$

che supera non solo il quantile 3.84 al 95% ma anche quello 6.63 al 99% e in effetti di molto, e quindi respinge l'ipotesi (nulla) di indipendenza del fumo dalla bronchite. Di per sé quest'affermazione non dice se il fumo sembra favorire o sfavorire la bronchite, ma un'attenta analisi dei dati ci porta a decidere per il favorire.

39.3 Sulle condizioni per la validità del test

Varie condizioni sulla non eccessiva "capricciosità" dei 4 valori vengono fatte dai vari Autori, per garantire la bontà della comunque inevitabile approssimazione del test. La questione è comunque contemporaneamente complessa e largamente arbitraria, e non la tratteremo. Confidiamo piuttosto che il software che il ricercatore poi userà nella pratica, dia un suo avvertimento in casi dubbi.

39.4 Test di Fischer esatto

Si può evitare l'approssimazione che è intrinseca nel test del chi quadrato d'indipendenza col *Test di Fischer esatto*, che ha una for-

mula molto più complicata che viene comunque calcolata in un'istante dai computer, a meno che i numeri non siano esorbitanti.

Il test di Fischer esatto è migliore del test del chi quadrato, che tuttavia è in generale ancora preferito *per tradizione*.

39.5 Morti per tutte le cause

E certo, una terza colonna, con i morti per tutte le cause, *all causes*, sarebbe utilissima per confronto. Per esempio così:

	malattia	non malattia	morti per tutte le cause
vaccino	a	b	c
placebo	d	e	f

Su trial clinici grandi e sufficientemente prolungati nel tempo, i morti ci saranno inevitabilmente, in entrambe le classi: come vago ordine di grandezza, si consideri che in un anno muore circa l'1% di un insieme di persone prese a caso in Stati sufficientemente sviluppati, il che corrisponde a circa 1 morto ogni 1000 persone in 1 mese di trial clinico. Leggiamo in

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8461570/> nell'articolo scientifico⁽¹⁹⁵⁾ *Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine through 6 Months* (2021) sul ciclopico trial clinico di lungo periodo (si arrivò a 6 mesi e poi lo sospesero – dissero – *per motivi etici*) del celeberrimo vaccino covid (2020) della Pfizer, il BNT162b2:

During the blinded, placebo-controlled period, 15 participants in the BNT162b2 group and 14 in the placebo group died

¹⁹⁵Thomas SJ, Moreira ED Jr, Kitchin N, Absalon J, Gurtman A, Lockhart S, Perez JL, Pérez Marc G, Polack FP, Zerbini C, Bailey R, Swanson KA, Xu X, Roychoudhury S, Koury K, Bouguermouh S, Kalina WV, Cooper D, Frenck RW Jr, Hammitt LL, Türeci Ö, Nell H, Schaefer A, Ünal S, Yang Q, Liberator P, Tresnan DB, Mather S, Dormitzer PR, Şahin U, Gruber WC, Jansen KU; C4591001 Clinical Trial Group. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine through 6 Months. *N Engl J Med.* 2021 Nov 4;385(19):1761-1773. doi: 10.1056/NEJMoa2110345. Epub 2021 Sep 15. PMID: 34525277; PMCID: PMC8461570.

39.6 Esempio di applicazione in un caso reale

Nel 2024–2025 l’Agenzia Europea dei Medicinali (EMA) ha avviato una revisione di sicurezza sul farmaco Oxbryta (voxelotor), utilizzato per il trattamento dell’anemia⁽¹⁹⁶⁾ falciforme. L’EMA ha⁽¹⁹⁷⁾ comunicato (2025) alcuni risultati preliminari provenienti da studi clinici ancora in corso, evidenziando un possibile segnale di incremento della mortalità nei bambini trattati con Oxbryta rispetto al placebo.

Secondo tali fonti, in uno studio pediatrico randomizzato in doppio cieco condotto in soggetti ad alto rischio di ictus, sono stati osservati 8 decessi nel gruppo trattato con Oxbryta e 2 nel gruppo placebo. Tuttavia, la documentazione pubblicamente disponibile non riporta con precisione il numero esatto di partecipanti per ciascun braccio, indicando soltanto una randomizzazione 1:1 su un totale di circa 236 bambini. Assumendo dunque una suddivisione approssimativa di 118 soggetti per gruppo – ipotesi ragionevole ma non verificabile con certezza in assenza del report completo dello studio – si ottiene la seguente tabella di contingenza:

	Deceduti	Vivi	Totale
Trattamento (Oxbryta)	8	110	118
Placebo	2	116	118
Totale	10	226	236

La condizione classica⁽¹⁹⁸⁾ per l’impiego dell’approssimazione χ^2 risulta soddisfatta ma proprio al limite. Tale caratteristica, unita al numero molto ridotto di eventi osservati (soltanto 10 decessi), suggerisce una certa cautela nell’interpretazione dell’approssimazione asintotica.

¹⁹⁶Sickle cell disease (SCD), una malattia genetica.

¹⁹⁷I documenti principali, liberamente consultabili, includono la comunicazione (2025) di EMA [EMA confirms suspension of Oxbryta](#) e la corrispondente nota (2025) di AIFA [EMA confirms suspension of Oxbryta \(press release\)](#).

¹⁹⁸Richiede che tutte le frequenze attese siano almeno pari a 5. Sotto l’ipotesi nulla di indipendenza tra trattamento e mortalità, le frequenze attese risultano pari rispettivamente a 5 decessi e 113 sopravvivenze in ciascun gruppo.

Applicando il test⁽¹⁹⁹⁾ del Chi quadrato (senza correzione di Yates) si ottiene un valore pari a $\chi^2 \approx 3.76$ con 1 grado di libertà, corrispondente a un valore di probabilità di circa $p \approx 0.053$. Questo risultato si colloca leggermente al di sopra della soglia tradizionale di significatività del 5%, configurando un esito statisticamente⁽²⁰⁰⁾ borderline.

È necessaria grande cautela, perché si tratta di questione grave, la di morti di bambini (non è il caso di un cosmetico risultato di dubbia efficacia per le rughe).

Il farmaco comunque – come leggiamo nel comunicato dell’EMA – è stato sospeso, in base a questo e a un altro studio (con 9 morti e altre condizioni negative rilevate).

39.7 Il solito problema: cosa misuriamo? Placebo

Con quale tipo di placebo fare il confronto?

In passato erano comuni trial clinici con placebo tutt’altro che inerti, per esempio un altro vaccino o una soluzione di idrossido di alluminio testando il vaccino HPV. Qua possiamo solo rimarcare lo sconcerto di taluni per questa pratica, senza poter approfondire. Ma nei recenti (2020) trial per i vaccini contro il covid-19 si è salutato con compiacimento l’uso della vera soluzione salina come placebo.

¹⁹⁹In situazioni caratterizzate da eventi rari e da celle con conteggi minimi, il test esatto di Fisher rappresenta un’alternativa metodologicamente più appropriata; applicato qui, esso produce un valore p bilaterale approssimativamente pari a 0.10, valore più conservativo e coerente con le proprietà del test in presenza di piccole numerosità campionarie.

²⁰⁰Va inoltre sottolineato che l’incertezza relativa alla distribuzione esatta dei partecipanti tra i due gruppi potrebbe modificare lievemente i valori numerici dei test, pur senza alterare la conclusione generale: entrambi i metodi suggeriscono l’esistenza di un possibile segnale di differenza nella mortalità, pur senza fornire evidenze statistiche conclusive al livello convenzionale del 5%. Questo caso reale mostra come la corretta interpretazione dei risultati debba integrare l’analisi statistica, le limitazioni dei dati disponibili e la rilevanza clinica degli eventi considerati, specialmente quando si tratta di sicurezza farmacologica in popolazioni pediatriche vulnerabili.

39.8 Trucchi del mestiere

E certo sarebbe una cosa carina che il doppio cieco non venga beffato con pillole... di diversi colori (vedi <https://www.theguardian.com/business/2014/apr/10/tamiflu-saga-drug-trials-big-pharma>

Ma non ci si illuda troppo. Nel trial clinico che ha portato all'autorizzazione in via emergenziale del vaccino contro il covid-19, la Pfizer ha presentato un testo impresentabile per uno statistico (eppure è stato presentato e approvato) in cui il numero di *exclusions* (casi esclusi per un qualche motivo: uno dei grossi problemi della Statistica nel mondo reale) era quintuplo nel gruppo vaccino rispetto al gruppo placebo, senza dare alcuna spiegazione plausibile per questo fatto, inverosimile dal punto di vista statistico: se veramente chi somministrava vaccini e placebo non sapeva, le *exclusions* avrebbero dovuto trovarsi bilanciatamente nei due gruppi. Si veda sul British Medical Journal <https://blogs.bmj.com/bmj/2021/01/04/peter-doshi-pfizer-and-modernas-95-effective-vaccines-we-need-more->

What is concerning is the imbalance between randomized groups in the number of these excluded individuals: 311 from the vaccine group vs 60 on placebo.

Ma le manipolazioni statistiche andavano ben oltre, si veda <https://blogs.bmj.com/bmj/2021/02/05/clarification-pfizer-and-modernas-95-effective-vaccines-we-need-more-> riducendosi la dichiarata efficacia del 95% a 2 possibilità

calculations of 19% and 29% vaccine efficacy

Ma poi, in definitiva, è saltato fuori che c'erano taroccamenti sui dati stessi: <https://www.bmj.com/content/375/bmj.n2635>, per esempio

vaccine packaging materials with trial participants' identification numbers written on them left out in the open, potentially unblinding participants

L'importante, comunque, è che il vaccino previene il contagio, la contagiosità, la malattia, la malattia severa, la terapia intensiva, e

la morte, nel mondo reale, al di là del trial clinico iniziale: <https://www.adnkronos.com/vaccini-moderna-e-pfizer-protezione-per-anni-dal-cov-3IkcyLUR4dzdvJPbVXvyL>

E, per buona misura, che la ricercatrice che ha documentato fotograficamente i tarocamenti sia stata licenziata. Queste cose non s'han da fare, perchè potrebbero minare la *fede nella Scienza*.

39.9 Duplice esempio da un articolo scientifico

In Medicina c'è una diatriba di lungo corso sull'ipotetica correlazione fra vermi intestinali e bruxismo (digrignamento dei denti nel sonno, in particolare nei bambini).

Un articolo scientifico⁽²⁰¹⁾ considera bambini di 3-6 anni e in 50 con bruxismo e 50 senza trova

Parasitic infections were observed in 19 (11/50 cases and 8/50 controls) children

cioè

	bruxism	without
intestinal parasitic infection	11	8
without	39	42

e conclude

The frequency of parasitic infections does not indicate any significantly greater frequency rate in the case group.

²⁰¹Tehrani MH, Pestechian N, Yousefi H, Sekhavati H, Attarzadeh H. The Correlation between Intestinal Parasitic Infections and Bruxism among 3-6 Year-Old Children in Isfahan. Dent Res J (Isfahan). 2010 Summer;7(2):51-5. PMID: 22013457; PMCID: PMC3177368.

Verifichiamolo con la Formula (156) del test del chi quadrato d'indipendenza:

$$\begin{aligned}
 T &:= \frac{(ad - bc)^2 (a + b + c + d)}{(a + b)(a + c)(b + d)(c + d)} = \\
 &= \frac{(11 \cdot 42 - 8 \cdot 39)^2 (11 + 8 + 39 + 42)}{(11 + 8)(11 + 39)(8 + 42)(39 + 42)} = \\
 &= \frac{(462 - 312)^2 \cdot 100}{19 \cdot 50 \cdot 50 \cdot 81} = \\
 &= \frac{150^2 \cdot 100}{3\,847\,500} = \\
 &= \frac{22\,500}{38\,475} \tag{157}
 \end{aligned}$$

che non è > 3.84 .

Potrebbero concludere così ma forse vogliono proprio dimostrare che il bruxismo è collegato ai parassiti, almeno a quelli più cattivi; distinguono i soggetti coi parassiti in soggetti con parassiti patogenici e soggetti con parassiti non patogenici:

However, since all the 11/50 of the detected infections in the case group included pathogenic parasites while from the 8/50 parasitic infections found in the controls, only 3/50 involved pathogenic infections, a statistically significant relationship can be established between parasitic infection and bruxism.

Cioè

	bruxism	without
intestinal pathogenic parasitic infection	11	3
without	39	47

Verifichiamolo con la Formula (156) del test del chi quadrato d'indipendenza:

$$\begin{aligned}
 T & ::= \frac{(ad - bc)^2 (a + b + c + d)}{(a + b)(a + c)(b + d)(c + d)} = \\
 & = \frac{(11 \cdot 47 - 3 \cdot 39)^2 (11 + 3 + 39 + 47)}{(11 + 3)(11 + 39)(3 + 47)(39 + 47)} = \\
 & = \dots \approx 5.316. > 3.84 \qquad (158)
 \end{aligned}$$

Perciò nell'abstract possono scrivere, con quasi bizantina sottigliezza.

Comparison for the frequency distribution of intestinal parasitic infections between the two groups was performed using the chi-square test (...) Parasitic infections were observed in 19 (11/50 cases and 8/50 controls) children. A statistically significant relationship was observed between infection with pathogenic parasites and bruxism ($P < 0.05$).

(Apparentemente sbagliato, vedi (157), ma è giusto, vedi (158), perché nella prima statistica considera i parassiti, nella seconda i parassiti patogenici).

Si faccia attenzione alle singole parole, gli Scienziati quando vogliono bizantineggiano.



Nota. Di questa Lezione 39 ci si dovrebbe aspettare che lo studente conosca a memoria l'unica formula (155) *contemporaneamente* numerata *et* riquadrata (che si estende su 5 righe), oltre a tutto l'inquadramento teorico della questione.