Immagine che contiene esterni, segnale, bottiglia

Descrizione generata automaticamente

****

***Metodi Matematici e statistici:***

***Test chi quadro per l’indipendenza***

***NucifoRa Nicolas***

***A.a. 2019/2020***



Immagine che contiene oggetto

Descrizione generata automaticamente

Indice

1. Test chi quadro per l’indipendenza
   1. Test di ipotesi
   2. Test Chi quadro
   3. Test Chi quadro per l’indipendenza
2. Soluzione Software
   1. Ambiente di sviluppo, linguaggio e librerie utilizzate
   2. Struttura del programma
   3. Soluzione al test chi quadro per l’indipendenza
3. Esempi e test
4. Conclusione
5. Bibliografia

##### 

##### **Test chi quadro per l’indipendenza**

##### **Test di ipotesi**

##### **Un test di ipotesi è utilizzato per verificare la bontà di un’ipotesi, ovvero un’affermazione che ha come oggetti eventi del mondo reale, come:**

##### **Decidere se una moneta, dopo un certo numero di lanci, è truccata o meno**

##### **Lo scopo di un test di ipotesi è quello di prendere una decisione (detta decisione statistica), ovvero accettare o rifiutare l’ipotesi.**

##### **Per effettuare un test di ipotesi, occorre avere i seguenti dati:**

##### **Un campione della popolazione su cui effettuare il test (X1, X2, …, Xn), di numerosità n**

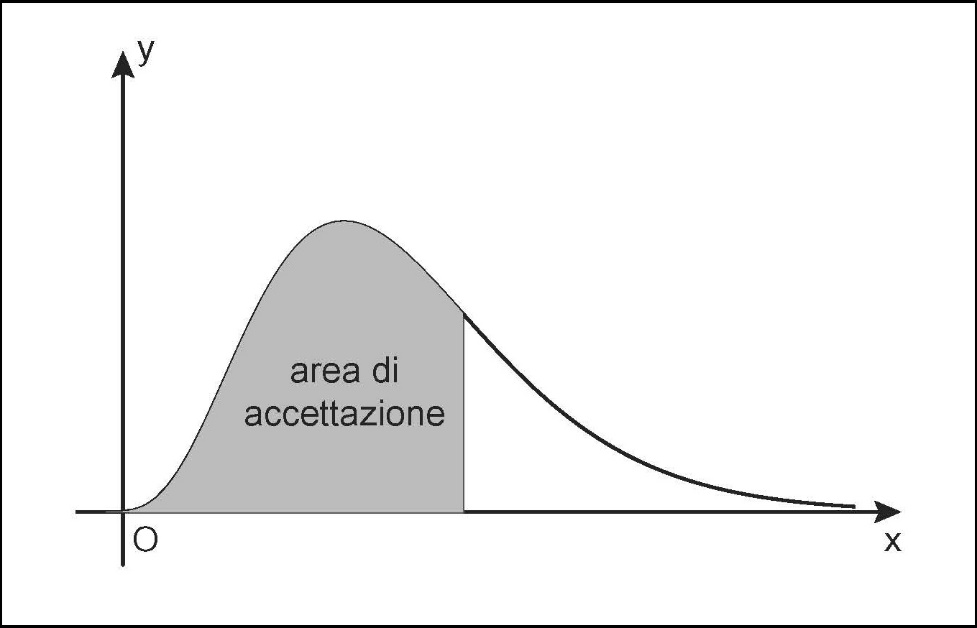
##### **Un’ipotesi nulla H0, che è quella da confermare**

##### **Un’ipotesi alternativa H1, che è l’ipotesi da considerare se** **H0 viene rifiutata**

##### **Una regione di accettazione, ovvero l’intervallo dei valori che ci permettono di accettare H0**

##### **Una regione critica, ovvero l’intervallo che ci porta a rifiutare H0**

##### **Un livello di significatività (margine d’errore) α, ovvero la probabilità d’errore che siamo disposti a commettere nell’affermare che H0 sia vera**

****

**Test X2 (chi quadro)**

I dati sono distribuiti secondo una X2

Dato un insieme di frequenze osservate (valori assunti dagli oggetti del campione durante il test), il test ci permette di determinare quanto questi valori si discostino dalle frequenze attese (i valori che dovrebbero assumere secondo la probabilità teorica).

Quindi definiamo:

* **n** = numerosità del campione
* **K** = numero delle classi (piccoli intervalli della possibile distribuzione che vogliamo testare)
* **Ok** = frequenze osservate, ovvero il n° di elementi del campione che cadono in **Ik** (intervallo k-esimo)
* **pk** = probabilità teorica del singolo campione Xi
* **Ak** = frequenze attese, dato dal prodotto **pk \* n**

Otteniamo la seguente tabella:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Campione | X1 | X2 | X3 | … | Xk |
| Osservate | O1 | O2 | O3 | … | Ok |
| Attese | A1 | A2 | A3 | … | Ak |

La variabile test X2 si ottiene:

Se H0 è vera e Ok≥5, si distribuisce secondo una chi quadro con K-1 gradi di libertà.

Fissato α, che assume valori molto piccoli (0.01; 0.05; 0.1), la regione critica è data dall’intervallo [X21-α, +∞[, quindi:

* Se W < X21-α : H0 viene accettata con probabilità di errore α
* Se W > X21-α: H0 viene rifiutata e viene considerata l’ipotesi alternativa H1

Oppure, è possibile calcolare il **p-value**, per evitare di effettuare il calcolo per ogni α.

**p-value = P(X > W) = = area della curva in [W, +∞[**

Se:

* P-value > α (per ogni α) : accettiamo H0
* P-value < α: rifiutiamo H0

**Test del X2 per l’indipendenza**

**Data una popolazione caratterizzata da due elementi X e Y, il test chi quadro per l’indipendenza ci permette di determinare se le due caratteristiche sono una indipendente dall’altra. Se ciò accade, possiamo dire che il verificarsi di una serie di eventi dipenda dal caso e non dalla natura dei dati sperimentali ottenuti.**

**Con questo tipo di test non è necessario conoscere la distribuzione del campione da studiare.**

**Per effettuare il test X2 per l‘indipendenza è necessario dare le seguenti definizioni:**

* **Phk = P(X=h & Y=k) con h=1,2,…,H & k=1,2,…,K (probabilità che X assuma H e Y assuma K)**
* **ph = P(X=h) = & pk = P(Y=k) = (Probabilità marginali: corrispondono alla somma delle righe e delle colonne della matrice ottenuta con le frequenze osservate X e Y)**

**Per verificare che le variabili sono una indipendente dall’altra:**

* **H0: Phk = P(X = h) \* P(Y = k) = ph \* pk**
* **H1: Phk ≠ ph \* pk**

**Definiamo, inoltre:**

* n = H + K (numero totali di campioni)
* Nhk : Numero di elementi per cui X = h & Y = k
* Nh:
* Mk:
* ph =
* pk =

**Le frequenze attese sono date da:**

**Essendo una matrice H x K, si può calcolare nel seguente modo:**

**Si calcola il p-value. Se il è maggiore di ogni α, allora le due variabili X e Y sono indipendenti**



**Soluzione Software**

**Ambiente di sviluppo**

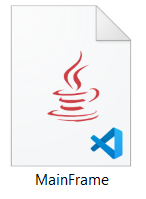
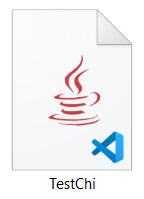
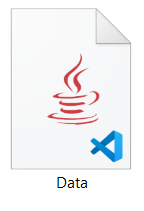
**L’ambiente di sviluppo utilizzato è “Netbeans”. La soluzione è stata implementata con il linguaggio java in quanto dispone di API (application programming interface) molto utili. Quelle utilizzate sono:**

* **Swing (per la creazione dell’interfaccia grafica)**
* **Apache common math 3.3 (per il calcolo del p-value)**

**Struttura software**

**Il programma è composto dalle seguenti classi:**

* **MainFrame: implementa l’interfaccia grafica del software. Contiene anche il main, per l’esecuzione del programma.**
* **Dato: implementa l’oggetto “dato”, ovvero i valori osservati**
* **TestChi: implementa il metodo per il calcolo del p-value, restituendo, inoltre, l’esito del test chi quadro per l’indipendenza**

****

**Soluzione al test chi quadro per l’indipendenza**

**L’interfaccia grafica del programma si presenta nel seguente modo:**

**Immagine che contiene screenshot

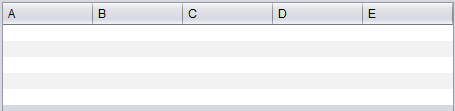
Descrizione generata automaticamente**

**Utilizzo:**

1. **Inserire la dimensione della matrice**



1. **Verrà creata una tabella NxM. Digitare i valori osservati (assicurarsi di premere invio dopo aver digitato l’ultimo elemento o non verrà inserito nella tabella, causando errori)**

****

1. **Seleziona il livello di significatività e premi “Calcola” per ottenere il risultato (selezionare alfa = 0.1 per determinare l’indipendenza totale)**



**Esempi e test**

**Da un indagine statistica svolta per sapere se gli uomini italiani svolgono le faccende domestiche, sono emersi i seguenti dati, classificati secondo l’età ed il numero di volte che in un mese vengono svolte le faccende domestiche:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Meno di 50 anni** | | **50-60 anni** | **Più di 60 anni** | |
| **Meno di 4 volte al mese** | | **25** | **15** | | **40** |
| **4 o più volte al mese** | | **65** | **15** | | **50** |

**Si può affermare che la frequenza con cui si svolgono le faccende domestiche sia indipendente dall’età?**

* **H0: la frequenza non dipende dall’età**
* **H1: la frequenza dipende dall’età**

**Immagine che contiene screenshot

Descrizione generata automaticamente**

**Possiamo affermare che la frequenza con cui si svolgono le faccende domestiche dipende dall’età dell’uomo (rifiutiamo H0).**

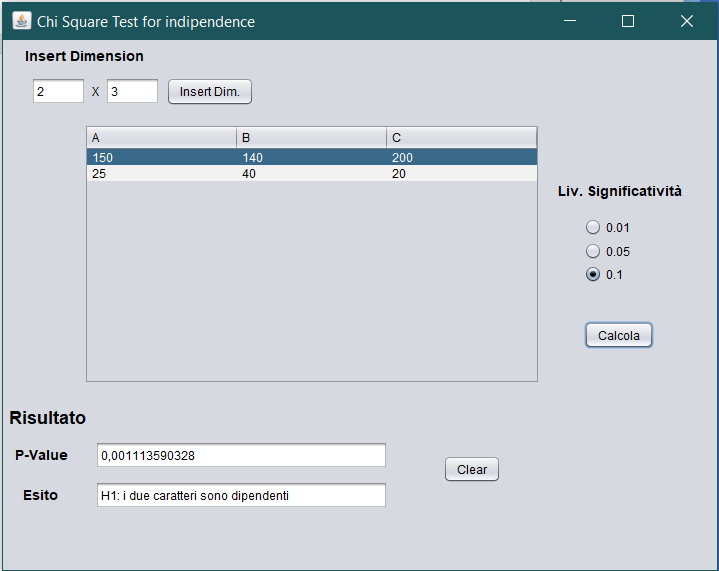
**Un ingegnere controlla la qualità dei pezzi prodotto da tre macchine diverse, ottenendo i seguenti dati:**

|  |  |  |
| --- | --- | --- |
| Macchina 1 | Macchina 2 | Macchina 3 |
| Buoni | 150 | 140 | 200 |
| Difettosi | 25 | 40 | 20 |

**La qualità dei pezzi difettosi dipende dalla macchina utilizzata?**

**H0: la qualità dei pezzi difettosi non dipendono dalla macchina che li ha prodotti**

**H1: la qualità dei pezzi difettosi dipende dalla macchina che li ha prodotti**

****

**Possiamo affermare che la qualità dei pezzi difettosi dipenda dalla macchina (rifiutiamo H0).**

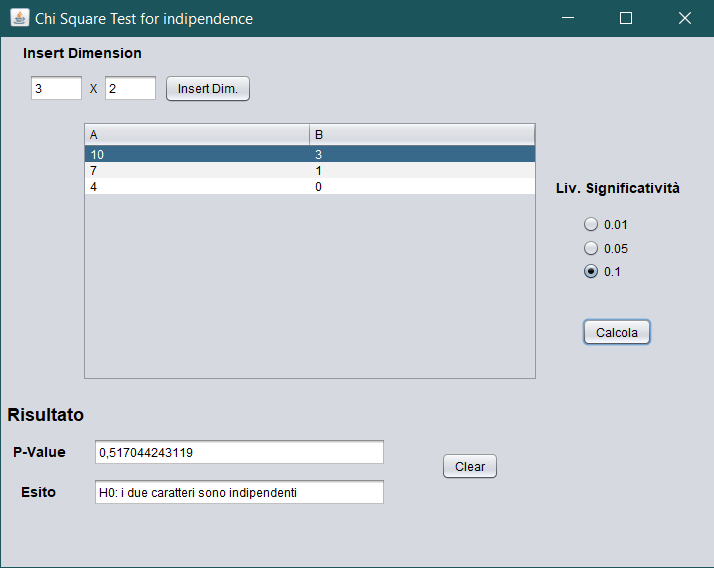
**In fase di sperimentazione clinica, un farmaco viene somministrato a 25 pazienti con tre dosaggi diversi, ottenendo i risultati riportati in tabella:**

|  |  |
| --- | --- |
| Guarigioni | Non guarigioni |
| Dosaggio A | 10 | 3 |
| Dosaggio B | 7 | 1 |
| Dosaggio C | 4 | 0 |

**Si può affermare che l’efficacia del farmaco non dipenda dal dosaggio?**

**H0: l’efficacia non dipende dal dosaggio**

**H1: l’efficacia dipende dal dosaggio**

****

**Possiamo affermare che l’efficacia del farmaco non dipenda dal dosaggio (accettiamo H0).**

**Conclusione**

**Lo sviluppo di tale software ha permesso di migliorare le mie abilità informatiche e ad accrescere il mio bagaglio culturale.**

**Il seguente progetto mi è stato proposto dal Prof. Muscato, che ringrazio tantissimo.**

**Bibliografia**

<http://www.dmi.unict.it/~muscato/MMStat.pdf> **(Appunti metodi matematici e statistici del prof. Muscato)**

<https://commons.apache.org/proper/commons-math/javadocs/api-3.1/org/apache/commons/math3/stat/inference/ChiSquareTest.html> **(Documentazione API Apache Common Math)**

Immagine che contiene esterni, segnale, bottiglia

Descrizione generata automaticamente