Analfabetismo statistico e decisioni informate

Matteo Paris

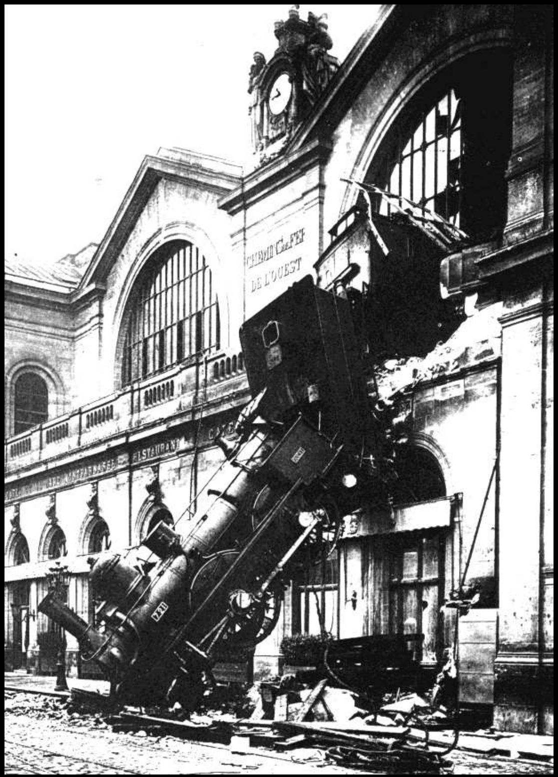
Dipartimento di Fisica – Università di Milano

matteo.paris@fisica.unimi.it

http://qinf.fisica.unimi.it/~paris

...a questo mondo non c’é niente di certo, a parte la morte e le tasse.

BENJAMIN FRANKLIN



Punti critici

Punti critici

- Frequenze e probabilità.

Punti critici

- Frequenze e probabilità.
- Percentuali e classi di riferimento.

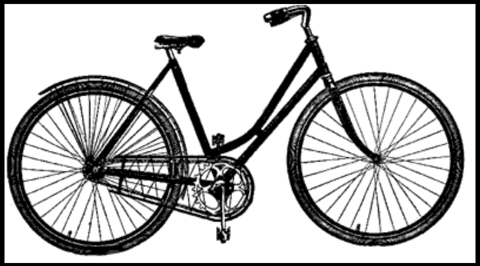
Punti critici

- Frequenze e probabilità.
- Percentuali e classi di riferimento.
- Probabilità condizionate.

Storiella

Storiella

Storiella



Storiella

Storiella

Il treno sarà in ritardo ?

Prendo la bicicletta ?

Decisioni in condizioni di incertezza !

Decisioni e statistica

Decisioni: studio, lavoro, previdenza, finanziamento della ricerca, gestione di fondi pubblici...

Decisioni e statistica

Decisioni: studio, lavoro, previdenza, finanziamento della ricerca, gestione di fondi pubblici...

- Percentuali, frequenze, probabilità condizionate

Decisioni e statistica

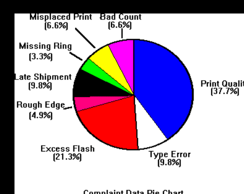
<u>Decisioni</u>: studio, lavoro, previdenza, finanziamento della ricerca, gestione di fondi pubblici...

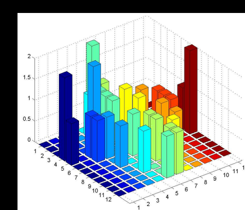
- Percentuali, frequenze, probabilità condizionate
- Rapporto rischio-beneficio, valutazione del rischio

Decisioni e statistica

Decisioni: studio, lavoro, previdenza, finanziamento della ricerca, gestione di fondi pubblici...

- Percentuali, frequenze, probabilità condizionate
- Rapporto rischio-beneficio, valutazione del rischio
- grafici, istogrammi

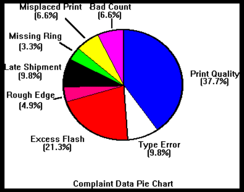


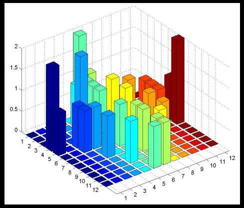


Decisioni e statistica

Decisioni: studio, lavoro, previdenza, finanziamento della ricerca, gestione di fondi pubblici...

- ▶ Percentuali, frequenze, probabilità condizionate
- ▶ Rapporto rischio-beneficio, valutazione del rischio
- ▶ grafici, istogrammi





É necessaria una cultura di base statistica
(nella popolazione e negli specialisti).

Tre frasi

- La mia cosa deve essere vernichiuta.

Tre frasi

- La mia cosa deve essere vernichiuta.
- Possiedo 14 macchine, ciascuna produce 3 camicie all'ora: dunque ogni ora sono prodotte $(14+3)=17$ camicie, che faccio quindi pagare 30 €.

Tre frasi

- La mia cosa deve essere vernichiuta.
- Possiedo 14 macchine, ciascuna produce 3 camicie all'ora: dunque ogni ora sono prodotte $(14 + 3) = 17$ camicie, che faccio quindi pagare 30 €.
- Le sue azioni hanno perso il 25$\%$ nel mese di Aprile, ma in Maggio hanno guadagnato il 30$\%$. Se mi ascolta la faró diventare ricco.

Due esempi

- Cosa significa 40% ? (test su 1000 tedeschi)

Due esempi

- Cosa significa 40% ? (test su 1000 tedeschi)

  ☐ un quarto ☐ 4 su 10 ☐ 1 su 40

Due esempi

- Cosa significa 40% ? (test su 1000 tedeschi)

  ☐ un quarto ☒ 4 su 10 ☐ 1 su 40

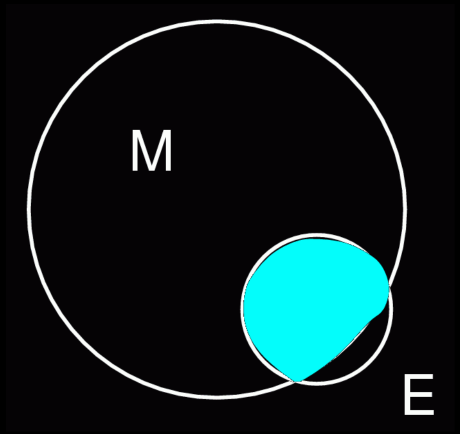
Due esempi

- Cosa significa 40% ? (test su 1000 tedeschi)

  ☐ un quarto ☒ 4 su 10 ☐ 1 su 40

- la maggioranza degli eroinomani consuma marijuana, <u>allora</u> la maggior parte di chi consuma marijuana consumerà eroina !

  (ministro degli interni bavarese)



Due esempi

- Cosa significa 40% ? (test su 1000 tedeschi)

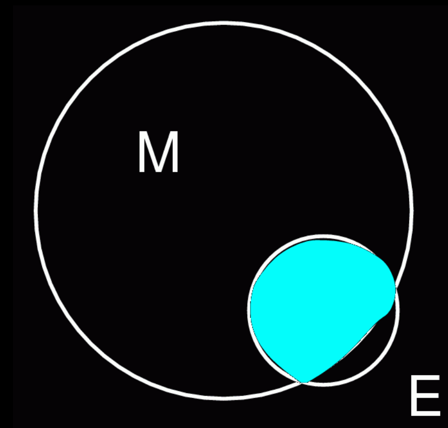
  ☐ un quarto ☒ 4 su 10 ☐ 1 su 40

- la maggioranza degli eroinomani

consuma marijuana, allora la maggior parte

di chi consuma marijuana consumerà eroina !

(ministro degli interni bavarese)



Non si puó basare un dibattito politico su un ragionamento errato (comunque la si pensi).

Perché questa lezione ?

Perché questa lezione ?

- I numeri sono linguaggio.

Perché questa lezione ?

- I numeri sono linguaggio.
- La fisica é una disciplina che non é nata ieri.

Perché questa lezione ?

- I numeri sono linguaggio.
- La fisica é una disciplina che non é nata ieri.
- La meccanica quantistica fa previsioni statistiche.

Variabili statistiche ed eventi

Variabili statistiche

- Variabile statistica X = esperimento con diversi possibili risultati (lancio di un dado, arrivo del treno, ...)

Variabili statistiche

- ► Variabile statistica X = esperimento con diversi possibili risultati (lancio di un dado, arrivo del treno, ...)

- ► Eventi x = possibili valori di una variabile statistica (x=1,...,6 o x=P,D per un dado, x=O,R per l'arrivo del treno,...)

Frequenza

Ripetiamo l'esperimento X un numero n di volte

Frequenza

Ripetiamo l'esperimento X un numero n di volte

▶ La frequenza $f(\mathsf{x})$ dell'evento x é la frazione di volte in cui esso accade

$$f(\mathsf{x}) = \frac{n_{\mathsf{x}}}{n}$$

Probabilità

Supponiamo di avere un modello per la variabile statistica (dado costruito con 6 facce eguali, capostazione che ha un solo giorno la settimana libero,...): i risultati possibili sono N, mentre i risultati favorevoli all'evento x sono N_x

Probabilità

Supponiamo di avere un modello per la variabile statistica (dado costruito con 6 facce eguali, capostazione che ha un solo giorno la settimana libero,...): i risultati possibili sono N, mentre i risultati favorevoli all'evento x sono N_x

► La probabilità $p(\mathrm{x})$ dell'evento x é la frazione di eventi favorevoli

$$p(\mathrm{x}) = \frac{N_{\mathrm{x}}}{N}$$

Legge dei grandi numeri

Legge dei grandi numeri

Quando facciamo un numero n grande di ripetizioni dell'esperimento X, la frequenza di un evento x converge alla probabilità

$$f(\mathsf{x}) \xrightarrow{n \gg 1} p(\mathsf{x})$$

Legge dei grandi numeri

Quando facciamo un numero n grande di ripetizioni dell'esperimento X, la frequenza di un evento x <u>converge</u> alla probabilità

$$f(\mathsf{x}) \xrightarrow{n \gg 1} p(\mathsf{x})$$

- Teorema di Bernoulli

Legge dei grandi numeri

Quando facciamo un numero n grande di ripetizioni dell'esperimento X, la frequenza di un evento x <u>converge</u> alla probabilità

$$f(\mathsf{x}) \xrightarrow{n \gg 1} p(\mathsf{x})$$

- Teorema di Bernoulli
- Teorema del limite centrale

Legge dei grandi numeri (2)

Quando non abbiamo un modello per la variabile statistica usiamo la frequenza <u>in tutti gli esperimenti precedenti</u> come indicatore della probabilità.

Legge dei grandi numeri (2)

Quando non abbiamo un modello per la variabile statistica usiamo la frequenza <u>in tutti gli esperimenti precedenti</u> come indicatore della probabilità.

► La distinzione tra probabilità e frequenza é spesso ignorata nel linguaggio comune.

Legge dei grandi numeri (2)

Quando non abbiamo un modello per la variabile statistica usiamo la frequenza in tutti gli esperimenti precedenti come indicatore della probabilità.

► La distinzione tra probabilità e frequenza é spesso ignorata nel linguaggio comune.

Esempi: Il treno ha il 10% di probabilità di essere in ritardo,
la probabilità di prendere il raffreddore a Dicembre é del 33$\%$.

Percentuali

- x ha probabilità 10% di accadere = su 100 ripetizioni dell’esperimento X mi aspetto che 10 diano risultato x

Percentuali

- x ha probabilità 10% di accadere = su 100 ripetizioni dell'esperimento X mi aspetto che 10 diano risultato x

É un modo comodo per confrontare risultati di osservazioni fatte con un numero di ripetizioni diverso.

Percentuali

- x ha probabilità 10% di accadere = su 100 ripetizioni dell'esperimento X mi aspetto che 10 diano risultato x

  É un modo comodo per confrontare risultati di osservazioni fatte con un numero di ripetizioni diverso.

- Puó essere fonte di ambiguità (qual'é il "100" ?)

Percentuali

- x ha probabilità 10% di accadere = su 100 ripetizioni dell'esperimento X mi aspetto che 10 diano risultato x

  É un modo comodo per confrontare risultati di osservazioni fatte con un numero di ripetizioni diverso.

- Puó essere fonte di ambiguità (qual'é il "100" ?)

- Le sue azioni hanno perso il 25% nel mese ☺ ☺ ☺
  di Aprile, ma in Maggio hanno guadagnato il 30%.
  Se mi ascolta....

Percentuali

- x ha probabilità 10% di accadere = su 100 ripetizioni dell’esperimento X mi aspetto che 10 diano risultato x

  É un modo comodo per confrontare risultati di osservazioni fatte con un numero di ripetizioni diverso.

- Puó essere fonte di ambiguità (qual’é il "100" ?)

- I suoi 100 milioni sono diventati 75 durante Aprile, ma poi durante Maggio sono diventati 97.5. ☹ ☹ ☹

  Se mi ascolta....

Percentuali (2)

Comunicato stampa: l'uso del farmaco Zaputin© riduce del 50% i decessi a causa della benzosi.

Percentuali (2)

Comunicato stampa: l'uso del farmaco Zaputin© riduce del 50% i decessi a causa della benzosi.

- Valutazione di un farmaco → prove a doppio cieco.

Percentuali (2)

<u>Comunicato stampa</u>: l'uso del farmaco Zaputin© riduce del 50% i decessi a causa della benzosi.

▶ Valutazione di un farmaco → prove a doppio cieco.

☺ I morti passano da 6 a 3 su un campione di 10 ?

Percentuali (2)

Comunicato stampa: l'uso del farmaco Zaputin© riduce del 50% i decessi a causa della benzosi.

- Valutazione di un farmaco → prove a doppio cieco.

☺ I morti passano da 6 a 3 su un campione di 10 ?

☹ I morti passano da 4 a 2 su un campione di 10^6 ?

Valutazione di un farmaco

La stessa riduzione relativa non corrisponde alla stessa riduzione assoluta. Cosa é piú rilevante per chi deve comprare il farmaco o finanziare nuove ricerche su di esso ?

Valutazione di un farmaco

La stessa riduzione relativa non corrisponde alla stessa riduzione assoluta. Cosa é piú rilevante per chi deve comprare il farmaco o finanziare nuove ricerche su di esso ?

- Riduzione rischio relativo

$$RRR = \frac{P - F}{P}$$

Caso1: RRR= 50%   Caso2: RRR=50%

Valutazione di un farmaco

La stessa riduzione relativa non corrisponde alla stessa riduzione assoluta. Cosa é piú rilevante per chi deve comprare il farmaco o finanziare nuove ricerche su di esso ?

- Riduzione rischio relativo

$$RRR = \frac{P - F}{P}$$

Caso1: RRR= 50% Caso2: RRR=50%

- Riduzione rischio assoluto (T=pazienti trattati)

$$RRA = \frac{P - F}{T}$$

Caso1: RRA= 30% Caso2: RRA=0.0002%

Valutazione di un farmaco

La stessa riduzione relativa non corrisponde alla stessa riduzione assoluta. Cosa é piú rilevante per chi deve comprare il farmaco o finanziare nuove ricerche su di esso ?

- Riduzione rischio relativo

$$RRR = \frac{P - F}{P}$$

Caso1: RRR= 50% Caso2: RRR=50%

- Riduzione rischio assoluto (T=pazienti trattati)

$$RRA = \frac{P - F}{T}$$

Caso1: RRA= 30% Caso2: RRA=0.0002%

- Cure necessarie per salvare una vita:

$$NNC = \frac{1}{RRA}$$

Caso1: NNC $\simeq$ 3 Caso2: NNC $\simeq$ 500000

Piú di una variabile statistica...

Probabilità congiunta

Probabilità congiunta

- Due variabili: X = altezza, Y = colore dei capelli
- Eventi: x = a,m,b   y = b,c,s

Probabilità congiunta

- Due variabili: X = altezza, Y = colore dei capelli
- Eventi: x = a,m,b   y = b,c,s
- Probabilità congiunta (x e y)

$$p(\mathrm{x}, \mathrm{y}) = \frac{n_{\mathrm{xy}}}{n} = p(\mathrm{y}, \mathrm{x})$$

Probabilità congiunta

- Due variabili: X = altezza, Y = colore dei capelli
- Eventi: x = a,m,b   y = b,c,s
- Probabilità congiunta (x e y)

$$p(\mathrm{x},\mathrm{y}) = \frac{n_{\mathrm{xy}}}{n} = p(\mathrm{y},\mathrm{x})$$

- E la probabilità totale di essere alti ?

$$p(\mathrm{a}) = \frac{n_{\mathrm{a}}}{n} = \frac{n_{\mathrm{ab}} + n_{\mathrm{ac}} + n_{\mathrm{as}}}{n} = p(\mathrm{a},\mathrm{b}) + p(\mathrm{a},\mathrm{c}) + p(\mathrm{a},\mathrm{s})$$

Probabilità condizionata

Quanti ragazzi alti <u>tra quelli castani</u> ?

Probabilità condizionata

Quanti ragazzi alti tra quelli castani ?

- Probabilità condizionata (x se y)

$$p(\mathrm{x}|\mathrm{y}) = \frac{n_{\mathrm{xy}}}{n_{\mathrm{y}}}$$

(notate che in questa definizione il "100" é diverso)

Teorema di Bayes

$$p(\mathsf{x},\mathsf{y}) = \frac{n_{\mathsf{xy}}}{n} = \frac{n_{\mathsf{xy}}}{n_{\mathsf{y}}}\,\frac{n_{\mathsf{y}}}{n} = p(\mathsf{x}|\mathsf{y})\,p(\mathsf{y})$$

Teorema di Bayes

$$p(\mathsf{x},\mathsf{y}) = \frac{n_{\mathsf{xy}}}{n} = \frac{n_{\mathsf{xy}}}{n_{\mathsf{y}}}\,\frac{n_{\mathsf{y}}}{n} = p(\mathsf{x}|\mathsf{y})\;p(\mathsf{y})$$

$$p(\mathsf{x},\mathsf{y}) = \frac{n_{\mathsf{xy}}}{n} = \frac{n_{\mathsf{xy}}}{n_{\mathsf{x}}}\,\frac{n_{\mathsf{x}}}{n} = p(\mathsf{y}|\mathsf{x})\;p(\mathsf{x})$$

Teorema di Bayes

$$p(\mathsf{x},\mathsf{y}) = \frac{n_{\mathsf{xy}}}{n} = \frac{n_{\mathsf{xy}}}{n_{\mathsf{y}}}\,\frac{n_{\mathsf{y}}}{n} = p(\mathsf{x}|\mathsf{y})\,p(\mathsf{y})$$

$$p(\mathsf{x},\mathsf{y}) = \frac{n_{\mathsf{xy}}}{n} = \frac{n_{\mathsf{xy}}}{n_{\mathsf{x}}}\,\frac{n_{\mathsf{x}}}{n} = p(\mathsf{y}|\mathsf{x})\,p(\mathsf{x})$$

- Teorema di Bayes

$$p(\mathsf{x}|\mathsf{y})\,p(\mathsf{y}) = p(\mathsf{y}|\mathsf{x})\,p(\mathsf{x})$$

Esempi in campo medico

Mammografie

Mammografie

- Una donna di 40 anni ha probabilità 1% di sviluppare un cancro al seno. Il test é affidabile, nel senso che in una donna malata il test risulta positivo nel 99% dei casi. Tra le donne sane il 9% avrà comunque il test positivo (probabilità di falso allarme).
Domanda: se una donna ha il test positivo deve preoccuparsi ? Ovvero, qual'e' la probabilità di essere malata per una donna con il test positivo ?

Mammografie

- Una donna di 40 anni ha probabilità 1% di sviluppare un cancro al seno. Il test é affidabile, nel senso che in una donna malata il test risulta positivo nel 99% dei casi. Tra le donne sane il 9% avrà comunque il test positivo (probabilità di falso allarme).
Domanda: se una donna ha il test positivo deve preoccuparsi ? Ovvero, qual'e' la probabilità di essere malata per una donna con il test positivo ?

- Su cento donne di 40 anni una ha il cancro al seno. Il suo mammogramma sará quasi sicuramente positivo. Delle rimanenti 99 donne, sane, circa 9 (il 9% di 99) avranno comunque il mammogramma positivo.
Domanda: tra le donne con il mammogramma positivo, quante hanno veramente il cancro al seno ?

Mammografie (2)

- Probabilità condizionata

$$p(\mathsf{m}|+) = \frac{p(+|\mathsf{m})p(\mathsf{m})}{p(+)}$$

Mammografie (2)

- Probabilità condizionata

$$p(\mathsf{m}|+) = \frac{p(+|\mathsf{m})p(\mathsf{m})}{p(+)}$$

- $p(\mathsf{m}) = 1/100$

Mammografie (2)

- Probabilità condizionata

$$p(\mathsf{m}|+) = \frac{p(+|\mathsf{m})p(\mathsf{m})}{p(+)}$$

- $p(\mathsf{m}) = 1/100$
- $p(+|\mathsf{m}) \simeq 1$

Mammografie (2)

- Probabilità condizionata

$$p(\mathsf{m}|+) = \frac{p(+|\mathsf{m})p(\mathsf{m})}{p(+)}$$

- $p(\mathsf{m}) = 1/100$
- $p(+|\mathsf{m}) \simeq 1$

- $$\begin{aligned} p(+) &= p(+|\mathsf{m})p(\mathsf{m}) + p(+|\mathsf{s})p(\mathsf{s}) \simeq p(\mathsf{m}) + p(+|\mathsf{s}) \\ &= 1/100 + 9/100 = 1/10 \end{aligned}$$

Mammografie (2)

- Probabilità condizionata $$p(\mathsf{m}|+) = \frac{p(+|\mathsf{m})p(\mathsf{m})}{p(+)}$$
- $p(\mathsf{m}) = 1/100$
- $p(+|\mathsf{m}) \simeq 1$
- $$p(+) = p(+|\mathsf{m})p(\mathsf{m}) + p(+|\mathsf{s})p(\mathsf{s}) \simeq p(\mathsf{m}) + p(+|\mathsf{s})$$ $$= 1/100 + 9/100 = 1/10$$
- $$p(\mathsf{m}|+) \simeq \frac{p(\mathsf{m})}{p(\mathsf{m}) + p(+|\mathsf{s})} = \frac{1/100}{1/10} = \frac{1}{10}$$

Mammografie (3)

Rapporto costo/beneficio dello screening mammografico

Mammografie (3)

Rapporto costo/beneficio dello screening mammografico

- Uno studio svedese, condotto su campioni per complessive 280.000 donne di 40 anni per dieci anni, ha rivelato che tra le donne sottoposte a screening (annuale) sono morte 3 donne su 1000 per cancro al seno, mentre tra le donne non sottoposte a screening sono morte 4 donne ogni 1000 per cancro al seno.

Mammografie (3)

Rapporto costo/beneficio dello screening mammografico

- Uno studio svedese, condotto su campioni per complessive 280.000 donne di 40 anni per dieci anni, ha rivelato che tra le donne sottoposte a screening (annuale) sono morte 3 donne su 1000 per cancro al seno, mentre tra le donne non sottoposte a screening sono morte 4 donne ogni 1000 per cancro al seno.

- RRR=25%
- RRA=0.1%
- NNC=1000

Mammografie (3)

Rapporto costo/beneficio dello screening mammografico

- Uno studio svedese, condotto su campioni per complessive 280.000 donne di 40 anni per dieci anni, ha rivelato che tra le donne sottoposte a screening (annuale) sono morte 3 donne su 1000 per cancro al seno, mentre tra le donne non sottoposte a screening sono morte 4 donne ogni 1000 per cancro al seno.

- RRR=25%
- RRA=0.1%
- NNC=1000

Come riportereste la notizia a chi deve decidere ?

Mammografie (3)

Rapporto costo/beneficio dello screening mammografico

▶ Uno studio svedese, condotto su campioni per complessive 280.000 donne di 40 anni per dieci anni, ha rivelato che tra le donne sottoposte a screening (annuale) sono morte 3 donne su 1000 per cancro al seno, mentre tra le donne non sottoposte a screening sono morte 4 donne ogni 1000 per cancro al seno.

▶ RRR=25% ▶ RRA=0.1% ▶ NNC=1000

Come riportereste la notizia a chi deve decidere ?

Come cittadini (contribuenti) non sareste piú tranquilli sapendo che chi decide mastica un po' di statistica ?

Test per l'HIV

"Un risultato positivo significa che nel suo sangue sono stati trovati gli anticorpi dell'HIV, e questo vuol dire che Lei ha un'infezione da HIV. É infetto per il resto della vita, e puó trasmettere HIV ad altre persone."

(ILLINOIS DPT OF PUBLIC HEALTH)

Test per l'HIV

"Un risultato positivo significa che nel suo sangue sono stati trovati gli anticorpi dell'HIV, e questo vuol dire che Lei ha un'infezione da HIV. É infetto per il resto della vita, e puó trasmettere HIV ad altre persone."

(ILLINOIS DPT OF PUBLIC HEALTH)

Nel decennio 1980-1990 in Florida, su 22 donatori di sangue a cui era stato comunicato che il test ELISA (il piú semplice per l'HIV) era risultato positivo 7 si sono suicidati.

Test per l'HIV (2)

I test(s) per l'HIV sono considerati molto sicuri: un paziente infetto viene rivelato nel 99.9% dei casi, mentre in un paziente sieronegativo il test é negativo nel 99.99% dei casi, ovvero la probabilità di falso allarme é <u>solo</u> dello 0.01%=1/10000.

Test per l'HIV (2)

I test(s) per l'HIV sono considerati molto sicuri: un paziente infetto viene rivelato nel 99.9% dei casi, mentre in un paziente sieronegativo il test é negativo nel 99.99% dei casi, ovvero la probabilità di falso allarme é <u>solo</u> dello 0.01%=1/10000.

se il test é positivo la probabiblità che si sbagli é una su diecimila $\rightarrow$ l'infezione c'é quasi sicuramente (SBAGLIATO !)

Test per l'HIV (2)

I test(s) per l'HIV sono considerati molto sicuri: un paziente infetto viene rivelato nel 99.9% dei casi, mentre in un paziente sieronegativo il test é negativo nel 99.99% dei casi, ovvero la probabilità di falso allarme é solo dello 0.01%=1/10000.

se il test é positivo la probabiblità che si sbagli é una su diecimila

$\longrightarrow$ l'infezione c'é quasi sicuramente (SBAGLIATO !)

▶ Probabilità condizionata

$$p(\mathrm{i}|+) = \frac{p(+|\mathrm{i})p(\mathrm{i})}{p(+)}$$

Test per l'HIV (2)

I test(s) per l'HIV sono considerati molto sicuri: un paziente infetto viene rivelato nel 99.9% dei casi, mentre in un paziente sieronegativo il test é negativo nel 99.99% dei casi, ovvero la probabilità di falso allarme é solo dello 0.01%=1/10000.

se il test é positivo la probabiblità che si sbagli é una su diecimila
$\rightarrow$ l'infezione c'é quasi sicuramente (SBAGLIATO !)

- Probabilità condizionata

$$p(\mathsf{i}|+) = \frac{p(+|\mathsf{i})p(\mathsf{i})}{p(+)}$$

- $p(+|\mathsf{i}) \simeq 1$

Test per l'HIV (2)

I test(s) per l'HIV sono considerati molto sicuri: un paziente infetto viene rivelato nel 99.9% dei casi, mentre in un paziente sieronegativo il test é negativo nel 99.99% dei casi, ovvero la probabilità di falso allarme é <u>solo</u> dello 0.01%=1/10000.

se il test é positivo la probabiblità che si sbagli é una su diecimila

$\longrightarrow$ l'infezione c'é quasi sicuramente (SBAGLIATO !)

- Probabilità condizionata

$$p(\mathsf{i}|+) = \frac{p(+|\mathsf{i})p(\mathsf{i})}{p(+)}$$

- $p(+|\mathsf{i}) \simeq 1$
- $p(\mathsf{i})$ dipende dalla categoria di persone

Test per l'HIV (2)

I test(s) per l'HIV sono considerati molto sicuri: un paziente infetto viene rivelato nel 99.9% dei casi, mentre in un paziente sieronegativo il test é negativo nel 99.99% dei casi, ovvero la probabilità di falso allarme é <u>solo</u> dello 0.01%=1/10000.

se il test é positivo la probabiblità che si sbagli é una su diecimila

$\rightarrow$ l'infezione c'é quasi sicuramente (SBAGLIATO !)

- Probabilità condizionata

$$p(\mathsf{i}|+) = \frac{p(+|\mathsf{i})p(\mathsf{i})}{p(+)}$$

- $p(+|\mathsf{i}) \simeq 1$
- $p(\mathsf{i})$ dipende dalla categoria di persone
- $p(+) = p(+|\mathsf{i})p(\mathsf{i}) + p(+|\mathsf{s})p(\mathsf{s}) \simeq p(\mathsf{i}) + p(+|\mathsf{s})$

Test per l'HIV (3)

- Probabilità condizionata

$$p(\mathsf{i}|+) = \frac{p(\mathsf{i})}{p(\mathsf{i}) + p(+|\mathsf{s})}$$

Test per l'HIV (3)

- Probabilità condizionata $$p(\mathsf{i}|+) = \frac{p(\mathsf{i})}{p(\mathsf{i}) + p(+|\mathsf{s})}$$

<u>individui non a rischio</u> $p(\mathsf{i}) = 0,01\%$

$$p(\mathsf{i}|+) = \frac{1/10000}{1/10000 + 1/10000} = \frac{1}{2} = 50\%$$

Test per l'HIV (3)

- Probabilità condizionata $$p(\text{i}|+) = \frac{p(\text{i})}{p(\text{i}) + p(+|\text{s})}$$

<u>individui non a rischio</u> $p(\text{i}) = 0,01\%$

$$p(\text{i}|+) = \frac{1/10000}{1/10000 + 1/10000} = \frac{1}{2} = 50\%$$

<u>individui a rischio</u> $p(\text{i}) = 1,5\%$

$$p(\text{i}|+) = \frac{15/1000}{15/1000 + 1/10000} = 0,99 = 99\%$$

Esempi in campo giuridico

Processi

Siete accusati di omicidio e vi stanno processando. Contro di voi un solo indizio: il vostro DNA collima con una traccia trovata sul corpo della vittima. Viene chiamato un perito.

Processi

Siete accusati di omicidio e vi stanno processando. Contro di voi un solo indizio: il vostro DNA collima con una traccia trovata sul corpo della vittima. Viene chiamato un perito.

► "la probabilità che questa concordanza sia casuale é una su centomila"

Processi

Siete accusati di omicidio e vi stanno processando. Contro di voi un solo indizio: il vostro DNA collima con una traccia trovata sul corpo della vittima. Viene chiamato un perito.

- "la probabilità che questa concordanza sia casuale é una su centomila"
- "ogni centomila persone ci sarà una concordanza"

Processi

Siete accusati di omicidio e vi stanno processando. Contro di voi un solo indizio: il vostro DNA collima con una traccia trovata sul corpo della vittima. Viene chiamato un perito.

- "la probabilità che questa concordanza sia casuale é una su centomila"
- "ogni centomila persone ci sarà una concordanza"

Dove si é consumato il delitto ?

La fallacia dell'accusatore

Un ipotetico processo per furto. Alcuni testimoni riferiscono che il ladro, che é stato visto scappare dal luogo del furto, era un uomo *alto piú di 1.70 m*, *biondo*, con *barba* e *baffi*. La persona sotto processo corrisponde a queste caratteristiche. L'accusa riporta le seguenti considerazioni: la probabilità di essere alti piú di 1.70 m é 1/10, quella di essere biondi circa 1/5, quella di avere la barba 1/200 e quella di avere i baffi 1/100.

La fallacia dell'accusatore

Un ipotetico processo per furto. Alcuni testimoni riferiscono che il ladro, che é stato visto scappare dal luogo del furto, era un uomo *alto piú di 1.70 m*, *biondo*, con *barba* e *baffi*. La persona sotto processo corrisponde a queste caratteristiche. L'accusa riporta le seguenti considerazioni: la probabilità di essere alti piú di 1.70 m é 1/10, quella di essere biondi circa 1/5, quella di avere la barba 1/200 e quella di avere i baffi 1/100.

- Probabilità globale concordanza

$$p(\text{abbb}) = n_{\text{abbb}}/n = \frac{1}{10}\,\frac{1}{5}\,\frac{1}{200}\,\frac{1}{100} = 1/10^6$$

La fallacia dell'accusatore

Un ipotetico processo per furto. Alcuni testimoni riferiscono che il ladro, che é stato visto scappare dal luogo del furto, era un uomo *alto piú di 1.70 m*, *biondo*, con *barba* e *baffi*. La persona sotto processo corrisponde a queste caratteristiche. L'accusa riporta le seguenti considerazioni: la probabilità di essere alti piú di 1.70 m é 1/10, quella di essere biondi circa 1/5, quella di avere la barba 1/200 e quella di avere i baffi 1/100.

- Probabilità globale concordanza

$$p(\text{abbb}) = n_{\text{abbb}}/n = \frac{1}{10}\,\frac{1}{5}\,\frac{1}{200}\,\frac{1}{100} = 1/10^6$$

- La fallacia dell'accusatore

$$p(C|\text{abbb}) = 1 - p(\text{abbb})$$

La fallacia dell'accusatore (2)

- Il buon giudice

$$p(C|\text{abbb}) = \frac{p(\text{abbb}|C)p(C)}{p(\text{abbb})}$$

La fallacia dell'accusatore (2)

- Il buon giudice $$p(C|\text{abbb}) = \frac{p(\text{abbb}|C)p(C)}{p(\text{abbb})}$$
- Testimoni affidabili $p(\text{abbb}|C) = 1$

La fallacia dell'accusatore (2)

- Il buon giudice $$p(C|\text{abbb}) = \frac{p(\text{abbb}|C)p(C)}{p(\text{abbb})}$$

- Testimoni affidabili $p(\text{abbb}|C) = 1$
- Probabilità a priori $p(C) = 1/n$

La fallacia dell'accusatore (2)

- Il buon giudice $$p(C|\text{abbb}) = \frac{p(\text{abbb}|C)p(C)}{p(\text{abbb})}$$
- Testimoni affidabili $p(\text{abbb}|C) = 1$
- Probabilità a priori $p(C) = 1/n$
- Probabilità di concordanza $p(\text{abbb}) = n_{\text{abbb}}/n$

La fallacia dell'accusatore (2)

- Il buon giudice $$p(C|\text{abbb}) = \frac{p(\text{abbb}|C)p(C)}{p(\text{abbb})}$$
- Testimoni affidabili $p(\text{abbb}|C) = 1$
- Probabilità a priori $p(C) = 1/n$
- Probabilità di concordanza $p(\text{abbb}) = n_{\text{abbb}}/n$
- Probabilità di colpevolezza data la concordanza

$$p(C|\text{abbb}) = \frac{1}{n_{\text{abbb}}}$$

Ancora due questioni

- Scelta del campione

Ancora due questioni

- Scelta del campione
- Probabilità composte

Conclusioni

- La comprensione del linguaggio statistico é necessaria per prendere decisioni razionali, sia da parte del pubblico, sia da parte delle istituzioni.

Conclusioni

- La comprensione del linguaggio statistico é necessaria per prendere decisioni razionali, sia da parte del pubblico, sia da parte delle istituzioni.
- Esiste un diffuso analfabetismo statistico. I punti piú critici riguardano l'uso delle percentuali e delle probabilità condizionate.

Conclusioni

- La comprensione del linguaggio statistico é necessaria per prendere decisioni razionali, sia da parte del pubblico, sia da parte delle istituzioni.
- Esiste un diffuso analfabetismo statistico. I punti piú critici riguardano l'uso delle percentuali e delle probabilità condizionate.
- L'analfabetismo statistico é sfruttato da piú parti per indurre a decisioni non correttamente informate.

Conclusioni

- La comprensione del linguaggio statistico é necessaria per prendere decisioni razionali, sia da parte del pubblico, sia da parte delle istituzioni.
- Esiste un diffuso analfabetismo statistico. I punti piú critici riguardano l'uso delle percentuali e delle probabilità condizionate.
- L'analfabetismo statistico é sfruttato da piú parti per indurre a decisioni non correttamente informate.
- É necessario un programma di alfabetizzazione statistica.

piccola bibliografia

M. Garetto, *STATISTICA: Lezioni ed esercizi*,
`http://www.dm.unito.it/quadernididattici/2001d.html`

G. Gigerenzer, *Quando i numeri ingannano*
(Raffaello Cortina, Milano, 2002).

D. Castelle, *La scienza del caso*, (Dedalo, Bari, 1998).

Materials for the History of Statistics
`http://www.york.ac.uk/depts/maths/histstat/`