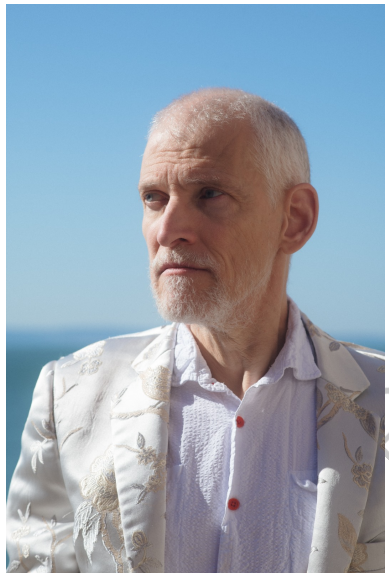
Le Compresse di Matematica per le Scienze della Farmacia

Lezioni 18-19

versione 29 ottobre 2026

BOZZA - DRAFT

Università degli Studi di Trieste



Alessandro Soranzo
Ricercatore universitario
Settore Scientifico Disciplinare:
MATH-03/A - Analisi matematica
Gruppo Scientifico Disciplinare:
01/MATH-03 - Analisi matematica, probabilità e statistica matematica

BOZZA - DRAFT

Indice

18 Introduzione alla Statistica Descrittiva **5**

18.1 Introduzione[10] 5

18.2 Riassumere i dati: quanti gatti e quanti cani avete?[14] 7

18.3 Esempio di possibile raccolta dati[18] 11

18.4 Il dataset e le operazioni su esso[22] 13

18.5 Le 4 più semplici statistiche di sintesi[26] 15

18.6 Esempio di farmacologia veterinaria[30] 16

18.7 Altre statistiche di sintesi[34] 17

18.8 Medie[38] 18

18.9 Inesistenza di una media migliore di tutte[42] 26

18.10 Disilludersi sulle medie anche senza outlier[46] 27

18.11 Stratificazione per età in Medicina e Farmacia[50] . . . 28

18.12 Sul trasformare dati ordinali in dati numerici[54] 29

18.13 Complementi – Illusioni percettive con le medie . . . 31

BOZZA - DRAFT

19 Bar chart, istogrammi, e altri diagrammi **37**

19.1 Diagrammi a torta[06] 37

19.2 Catastrofe Statistica: dati mancanti e drop out[10] . . . 39

19.3 Istogrammi a barre o bar chart, e istogrammi[14] 40

19.4 Frequenze assolute, relative, percentuali, cumulate[18] . 43

19.5 Avvertenze sui bar chart[22] 46

19.6 Eventuali asimmetrie nei dataset: skewness[26] 49

19.7 Funzioni a campana varie[30] 50

19.7.1 Esercizio risolto a – Bar chart[34] 55

19.7.2 Esercizio risolto b – Istogramma[38] 59

BOZZA - DRAFT

18 Introduzione alla Statistica Descrittiva

18.1 Introduzione[10]

La Statistica Descrittiva opera su dati.
Per avere i *dati*, è necessaria la *raccolta dati*, operazione preliminare – in generale complessa e fonte di innumerevoli errori e distorsioni – che tecnicamente non fa parte della Statistica Descrittiva.
Accenneremo alla raccolta dati, e cosa molto bella è fare esperienze in aula.

La Statistica Descrittiva si occupa essenzialmente di:

- **riassumere dati, tipicamente molti,**
  (se sono pochi può essere meglio elencarli)
- **presentare il riassunto,**
  il più comprensibilmente possibile: medie, diagrammi, e altro.

BOZZA - DRAFT

È cosa ben diversa, più semplice e storicamente antica, dalla Statistica Inferenziale che verrà trattata fra le Matematiche dell'Incertezza.

In una trattazione elementare della Statistica Descrittiva non si distingue fra *popolazione* e *campione*: abbiamo i dati che abbiamo, e quelli consideriamo popolazione, anche se in effetti sono un campione di una popolazione più ampia. La distinzione diventa invece essenziale nella Statistica Inferenziale.

In questo testo elementare, verrà tralasciata la questione della rilevazione materiale dei dati – con tutta la sua problematicità.

E in questa lezione **verrà trattata solo la sintesi** dei dati, ovvero la rappresentazione dei dati in una forma umanamente comprensibile e *trattabile*, cosa particolarmente utile se i dati sono più di una dozzina.

BOZZA - DRAFT

Si vuole riassumere i dati con 1 diagramma oppure 1 o pochi valori, per

comprenderli

confrontarli ← cosa fondamentale per i farmaci

divulgarli

e, a livello eticamente discutibile, presentarli manipolativamente (cosa che viene fatta amplissimamente: un fenomeno in *diminuzione* si trova sempre come presentarlo, almeno a livello di titolo riassuntivo, come *aumento*, se si vuole, usando artifici statistici formalmente legittimi).

BOZZA - DRAFT

18.2 Riassumere i dati: quanti gatti e quanti cani avete?[14]

Dati rilevati a Farmacia a Trieste il 17/11/2023, per gatti e cani:

```
G: 0 0 0 1 0 0 0 0 0 1 0 1 1 0 1 0 0 0 0 1 0 0 0 0 2 3 2 0 1 0
C: 1 1 0 1 0 2 1 0 0 0 0 0 0 2 4 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
```

Gli studenti di Farmacia partecipanti alla rilevazione
hanno da 0 a 3 gatti;
hanno da 0 a 4 cani.

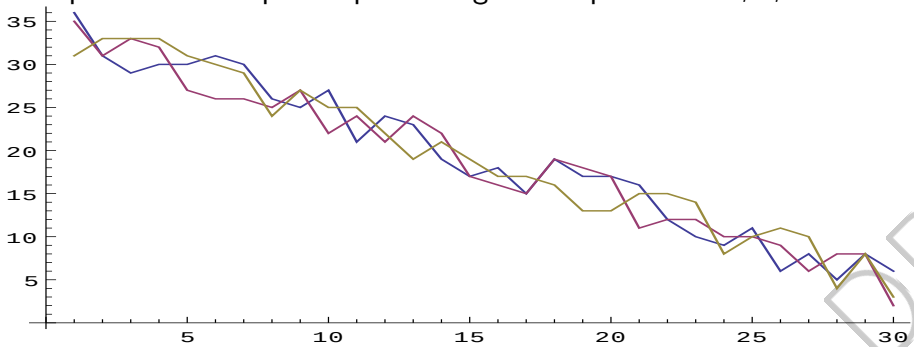
Si tratta delle statistiche del minimo e del massimo, di calcolo abbastanza semplice, eppure già molto espressive della situazione, di non così immediata comprensione semplicemente guardando il dataset, specialmente se avesse 100 numeri, come in questo esempio ipotetico:

2, 0, 3, 0, 3, 2, 1, 1, 2, 0, 2, 2, 2, 1, 1, 2, 1, 0, 2, 0, 3, 4, 0, 1, 4, 1, 2, 3, 0, 1, 0, 2, 1, 3, 3, 2, 0, 1, 0, 2, 0, 2, 3, 2, 1, 1, 2, 3, 2, 2, 4, 2, 1, 3, 3, 3, 2, 0, 2, 3, 0, 1, 2, 2, 2, 1, 4, 0, 2, 0, 0, 1, 2, 2, 1, 1, 1, 3, 0, 4, 0, 1, 2, 2, 1, 1, 2, 4, 1, 2, 3, 4, 2, 0, 0, 3, 2, 2, 3, 2

E con più calcoli potremmo trovare il numero medio di gatti, e di cani.
E vedremo molte altre forme di Statistica Descrittiva.

BOZZA - DRAFT

Esempio. Consideriamo un fenomeno in evoluzione da un giorno 1 a un giorno 30, tanto più negativo per la società quanto più sono grandi 3 parametri a, b, c.



Un *ingenuo* direbbe che la situazione sta migliorando: i parametri negativi diminuiscono. Invece i media del posto scrivono, a beneficio del cosiddetto *popolo bue*:

Giorno 2 Aumenta il tal parametro!
Giorno 3 Aumenta il tal parametro!
Giorno 4 Aumenta il tal parametro!
Giorno 5 Il Governo valuta la preoccupante situazione
Giorno 6 Aumenta il tal parametro!
Giorno 7 Il Governo valuta la preoccupante situazione
Giorno 8 Il Governo valuta la preoccupante situazione
Giorno 9 Aumenta il tal parametro!
Giorno 10 Aumenta il tal parametro!
Giorno 11 Aumenta il tal parametro!
Giorno 12 Aumenta il tal parametro!

BOZZA - DRAFT

Giorno 13 Aumenta il tal parametro!
Giorno 14 Aumenta il tal parametro!
Giorno 15 Il Governo valuta la preoccupante situazione
Giorno 16 Aumenta il tal parametro!
Giorno 17 Il Governo valuta la preoccupante situazione
Giorno 18 Aumenta il tal parametro!
Giorno 19 Il Governo valuta la preoccupante situazione
Giorno 20 Il Governo valuta la preoccupante situazione
Giorno 21 Aumenta il tal parametro!
Giorno 22 Aumenta il tal parametro!
Giorno 23 Il Governo valuta la preoccupante situazione
Giorno 24 Il Governo valuta la preoccupante situazione
Giorno 25 Aumenta il tal parametro!
Giorno 26 Aumenta il tal parametro!
Giorno 27 Aumenta il tal parametro!
Giorno 28 Aumenta il tal parametro!
Giorno 29 Aumenta il tal parametro!

Giorno 30 Il Governo valuta la preoccupante situazione

BOZZA - DRAFT

Avendo sentito nei vari giorni solo "aumenta", la generalità degli ascoltatori si forma l'idea che il fenomeno sia in crescita. Anche con trucchi più raffinati vengono fatti credere in aumento fenomeni che sono in diminuzione, a seconda degli interessi di chi produce informazione.

18.3 Esempio di possibile raccolta dati[18]

Fare in aula una raccolta dati è interessante ma bisogna fare estrema attenzione al rispetto per la persona (non si possono chiedere peso e altezza, nemmeno in questionari anonimi) e alla possibilità concreta che gli studenti conoscano il dato richiesto (non si può chiedere quanta vitamina D o rame hanno nel sangue perché in generale non lo sanno). La raccolta dati e successive analisi statistiche sono estremamente valevoli didatticamente e piacciono molto agli Studenti. Alcune possono essere fatte a voce, altre anonimamente.



Ecco una possibile raccolta dati, fattibile addirittura pubblicamente a voce in aula: Quanti animali domestici possiedi? Oppure si può dare un foglio a testa e nei 4 quadranti ideali del foglio scrivere i dati relativi a cani, gatti, roditori, uccellini, e al centro altri animali domestici.

BOZZA - DRAFT

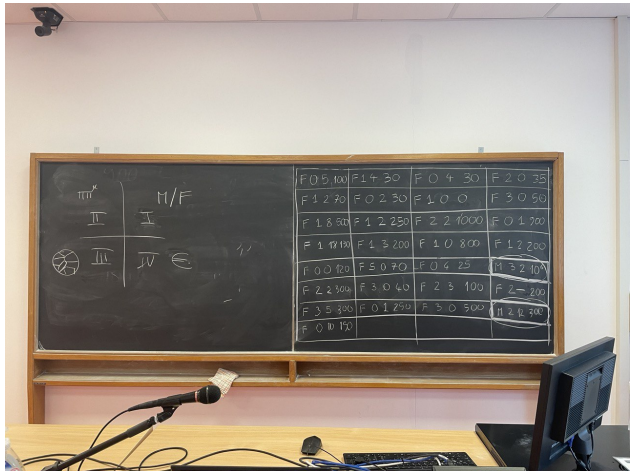


Figura 1: Raccolta dati, Matematica per Farmacia a Trieste, in un anno e con 4 domande che non specifichiamo. Si era consegnato un foglio bianco ad ogni presente in aula e le risposte andavano scritte nei 4 quadranti ideali del foglio. Raccolta dati anonima, risposte lette via via ad alta voce da una volontaria, e trascritte alla lavagna con successiva distruzione dei fogli. Il valore anomalo (*outlier*) un milione è trascritto come 10^6. Un trattino orizzontale indica un dato mancante: un partecipante ha scritto che non sa invece di rispondere con un numero, e forse intendeva 0 o proprio che non sa rispondere, in ogni caso è un dato mancante classificabile come "non sa/non risponde" oppure come dato mancante per qualsiasi causa.

BOZZA - DRAFT

18.4 Il dataset e le operazioni su esso[22]

L'insieme dei dati $x_1, ..., x_n$ si chiama *dataset* e non è un insieme in senso matematico perché le ripetizioni dei valori sono ammesse.

Si considerano dataset, via via più trattabili matematicamente:

• *nominali*: per esempio con valori in { pari, dispari }
pari, pari, dispari, pari, dispari, dispari (è un dataset)
oppure con valori in { integro, spezzato, bifido }
integro, spezzato, spezzato, bifido, integro, integro (è un dataset)

• *ordinali*: per esempio, per valutare un farmaco antidepressivo potremmo considerare questi valori non numerici
morte
grande peggioramento
medio peggioramento
lieve peggioramento
stabile
lieve miglioramento
medio miglioramento
grande miglioramento

BOZZA - DRAFT

e dopo il *trial clinico* avere un dataset di – per esempio – 172 valori di quel tipo.

Oppure, altro caso:

sono completamente d'accordo,
sono d'accordo,
sono parzialmente d'accordo,
non sono d'accordo,

non sono assolutamente d'accordo.

- *numerici*: 3, 1, 2, 4, 6, 1000, 6. Quelli che più ci interessano.

Anche nei dataset ordinali si potrà fissare una corrispondenza con numeri, ma non avrà alcun senso fare la media fra 3 soggetti con: lieve miglioramento, medio miglioramento, morte. Se invece la "perdita infinita" viene esclusa, una media in qualche modo si potrà anche fare, con opportuna trasformazione in valori numerici. (E alla fin fine – di necessità in virtù – qualcosa riusciranno a fare anche con la perdita infinita).

BOZZA - DRAFT

Fra le variabili nominali si distinguono quelle *dicotomiche* ovvero *binarie*, con 2 soli valori, come 0/1 vivo/morto, successo/fallimento.

Considereremo 2 capitoli della Statistica Descrittiva:
◇ le rappresentazioni grafiche: Lezioni 20 e 21;
◇ le statistiche di sintesi, funzioni $f(x_1, ..., x_n)$ con un numero n a priori indeterminato di argomenti. Inizieremo da queste.

BOZZA - DRAFT

18.5 Le 4 più semplici statistiche di sintesi[26]

Ecco alcune statistiche di sintesi per dati univariati, cioè per un dataset costituito da una singola lista $x_1, \ldots, x_n$:

- la numerosità: n ← per ogni dataset
- il minimo: $\min(x_1, \ldots, x_n)$ ← per un dataset almeno ordinale
- il massimo: $\max(x_1, \ldots, x_n)$ ← per un dataset almeno ordinale
- la somma: $x_1 + \cdots + x_n$ ← per un dataset numerico

18.6 Esempio di farmacologia veterinaria[30]

È immensamente più significativo dire che fra $100\,000$ ostriche sono state trovate perle con un minimo di 0 e un massimo di 4 in ogni esemplare per un totale di $41\,320$ perle, piuttosto che elencare il numero di perle 0, 0, 1, 0, 2, 0, 0, 0, 0, 1... di ciascuno degli esemplari. Abbiamo riassunto $100\,000$ numeri con 4 numeri:

numerosità	$100\,000$
minimo	0
massimo	4
somma	$41\,320$

e capiamo perfino meglio la situazione che con pagine di numeri!
E soprattutto possiamo meglio *confrontarla* con la situazione di un'altra vasca di ostriche, magari nutrite/trattate diversamente, che analogamente riassumeremo con quei 4 indici.

BOZZA - DRAFT

18.7 Altre statistiche di sintesi[34]

Per **dati univariati** cioè liste di numeri tratteremo:

- gli **indici di posizione**:
  - quelli che vorrebbero riassumere in 1 solo valore "medio" il complesso dei dati, e saranno l'argomento di questa Lezione 18;
  - quelli che vorrebbero riassumere in 5 valori, o corrispondentemente in 1 diagramma, il complesso dei dati, Lezione **??**;
- gli **indici di dispersione ovvero variabilità**, Lezione **??**, che vorrebbero quantificare con 1 numero la non omogeneità di dati numerici;
- **un indice di asimmetria** (*skewness*).

Esistono poi altri indici (per esempio la curtosi) che non tratteremo.

BOZZA - DRAFT

18.8 Medie[38]

Media (aritmetica). Da ora consideriamo un dataset $x_1, ..., x_n$.

$$M(x_1, ..., x_n) := \frac{x_1 + x_2 + ... + x_n}{n}$$

Esempio.
Sul covid-19 in Italia, leggiamo in
https://www.epicentro.iss.it/coronavirus/bollettino/Report-COVID-2019_4_ottobre.pdf

> Dati al 7 settembre 2020 (...) da 4190 deceduti per i quali è stato possibile analizzare le cartelle cliniche. Il numero medio di patologie osservate in questa popolazione è di 3,4

Facile, ottimo anche per i voti, ma il reddito medio di questi

$$3, 1, 2, 4, 6, 1000, 6$$

è 146, non poi così espressivo della situazione globale, a causa del valore 1000, un *outlier* ovvero valore anomalo. Tratteremo in seguito la questione dei valori anomali ma anticipiamo che talvolta vengono *eliminati*.

BOZZA - DRAFT

Media geometrica. Per n numeri positivi (invece la media aritmetica non lo richiede), è la radice n-esima del loro prodotto:

$$GM(x_1, ..., x_n) := (x_1 \cdot ... \cdot x_n)^{\frac{1}{n}} .$$

Nell'esempio numerico soprastante ≈ 7.05; con WolframAlpha

`geometric mean 3,1,2,4,6,1000,6`

Nota 1. Per quanto possa interessare il lettore, l'opinione dello scrivente è che più della metà delle volte in cui negli articoli scientifici si fa una media aritmetica, la media geometrica sarebbe stata più pertinente; e plausibilmente, in progresso di tempo la media geometrica sostituirà largamente la media aritmetica. Certo, la media geometrica è meno semplice calcolarla, almeno a mano.

Nota 2. Alcuni Autori ritengono che la media geometrica sia meno sensibile agli outlier della media aritmetica, ma questo non può essere affermato con certezza matematica, dipende da ogni singolo dataset.

BOZZA - DRAFT

Riguardo alle applicazioni non lontane dalle Scienze Biomediche, leggiamo in Wikipedia, l'enciclopedia libera, alla voce "Geometric mean":

> starting from 2010 the United Nations Human Development Index did switch to this mode of calculation, on the grounds that it better reflected the non-substitutable nature of the statistics being compiled and compared:
>
> > The geometric mean decreases the level of substitutability between dimensions [being compared] and at the same time ensures that a 1 percent decline in say life expectancy at birth has the same impact on the HDI as a 1 percent decline in education or income. Thus, as a basis for comparisons of achievements, this method is also more respectful of the intrinsic differences across the dimensions than a simple average.

BOZZA - DRAFT

In pratica, se un Paese porta l'aspettativa di vita da 40 a 60 anni, questo verrà ben rilevato facendo la media geometrica con il reddito pro-capite; se invece si facesse la media aritmetica, l'incremento di reddito da $4\,000$ a $4\,040$ dollari peserebbe di più: l'incremento è di 40 invece che 20, ma è solo l'1%, mentre da 40 a 60 è del 50%. La media geometrica appare migliore quando dobbiamo riassumere dati che variano su scale molto diverse. E anche in altri casi.

Media (aritmetica) ponderata (o pesata). Dati dei *pesi* $a_1, ..., a_n$, di somma 1, la media pesata del dataset $x_1, ..., x_n$ è $a_1x_1 + ... + a_nx_n$.
(Detto semplificatamente) pesi a_k maggiori di $\frac{1}{n}$ sono associati a dati che si vuol far "pesare" di più nella considerazione finale.

Per esempio nella media pesata

$$0.4\,voto_Matematica + 0.4\,voto_Chimica + 0.2\,voto_Marketing$$

il voto nell'esame di Marketing pesa/conta metà di ciascuno degli altri voti.

Mediana. È il numero centrale dei dati riordinati, adesso 4: $1, 2, 3, \underline{4}, 6, 6, 1000$. E se i dati sono in numero pari, si considera la media dei 2 centrali. La mediana è definita anche per valori *ordinali* almeno se il numero di dati è dispari.

WolframAlpha calcola online
la media con `mean`: link->
la mediana con `median`: link->
la media geometrica con `geometricmean`: link->
in tutti i casi facendo seguire la lista di numeri.

Leggiamo (11 novembre 2022) in https://www.epicentro.iss.it/coronavirus/sars-cov-2-decessi-italia

> al 10 gennaio 2022 (...) Complessivamente, le donne decedute dopo aver contratto infezione da SARS-CoV-2 hanno un'età più alta rispetto agli uomini (età mediane: donne 85 anni – uomini 80 anni).

BOZZA - DRAFT

Moda. È il valore più frequente. Nel nostro esempio è 6 ma in generale non è definita perché i numeri sono tutti diversi o perché 2 o più valori sono ripetuti ugualmente. Ecco per esempio un dataset *bimodale*: $6, 6, 6, 6, 7, 7.5, 7.5, 8, 8, 8, 8, 8.5, 8.5, 9, 9.5, 10, 10$.
La moda è definita anche per dati *nominali*, neppure ordinali, per esempio in ogni nazione c'è un cognome più frequente.

Confronto fra media, mediana e moda.

Su Wikipedia, l'enciclopedia libera, alla voce "Mode (statistics)" troviamo questo significativo esempio:

1, 2, 2, 3, 4, 7, 9; media=4, mediana=3, moda=2.

Leggiamo (11 novembre 2022) in https://www.epicentro.iss.it/coronavirus/sars-cov-2-decessi-italia

> al 10 gennaio 2022 (...) L'età media dei pazienti deceduti e positivi a SARS-CoV-2 è 80 anni (mediana 82 (...)).

BOZZA - DRAFT

Non ordinamento fisso di media, moda e mediana

Media, mediana e moda sono 3 valori che possono presentarsi in qualunque ordine fra loro: dipende dal dataset.

Media interquartile. Leggiamo su Wikipedia, l'enciclopedia libera:

> is best explained with an example. Consider the following dataset:
> 5, 8, 4, 38, 8, 6, 9, 7, 7, 3, 1, 6
> First sort the list from lowest-to-highest:
> 1, 3, 4, 5, 6, 6, 7, 7, 8, 8, 9, 38
> There are 12 observations (datapoints) in the dataset, thus we have 4 quartiles of 3 numbers. Discard the lowest and the highest 3 values:
> ~~1, 3, 4,~~ 5, 6, 6, 7, 7, 8, ~~8, 9, 38~~
> We now have 6 of the 12 observations remaining; next, we calculate the arithmetic mean of these numbers:
> xIQM = (5 + 6 + 6 + 7 + 7 + 8) / 6 = 6.5
> This is the interquartile mean.
> For comparison, the arithmetic mean of the original dataset is
> (5 + 8 + 4 + 38 + 8 + 6 + 9 + 7 + 7 + 3 + 1 + 6) / 12 = 8.5
> due to the strong influence of the outlier, 38.

Detto grezzamente, consideriamo il reddito medio delle persone medie, al netto di mendicanti e ricconi, che quelli ci sono comunque dappertutto.

Col metodo qua illustrato si trova la media interquartile per un numero di dati che è quadruplo; con artifici si definisce la media interquartile anche per un numero di dati non quadruplo ma non ce ne occupiamo.

Nota. Che 38 vada considerato outlier è opinabile.

BOZZA - DRAFT

ESERCIZIO$_{\mu 2019}$

$\approx$ Il carbonio risulta avere – almeno secondo alcuni Autori: non possiamo qua garantirlo in forma assoluta e fare Chimica o Fisica; si veda Wikipedia in inglese alla voce *Isotopes of carbon* – 12 isotopi non reperibili in natura (oltre a 3 reperibili in natura) con queste emivite approssimative:

$3,5 \times 10^{-21}$ s, 126,5 ms, 19,3 s, 20,364 min, 2,45 s, 0,747 s,
193 ms, 92 ms, 46,2 ms, 16 ms, $<$30 ns, 6,2 ms.

Dopo aver convertito minuti, millisecondi e nanosecondi in secondi con le note formule

1 min=60 s, 1 ms=0,001 s, 1 ns=$0,000\,000\,001$ s,

determinare la media interquartile delle emivite.

SVOLGIMENTO

Con la conversione in secondi le emivite sono, nell'ordine iniziale dei dati,

$3,5 \times 10^{-21}$ s, $0{,}126\,5$ s, 19,3 s, $1\,221,84$ s, 2,45 s, 0,747 s,
0,193 s, 0,092 s, $0,046\,2$ s, 0,016 s, $< 0,000\,000\,030$ s, $0,006\,2$ s.

Ovvero, in ordine crescente, omettendo l'unità di misura,

$3,5 \times 10^{-21} \rightarrow$ questo e il seguente potrebbero doversi scambiare
$< 0,000\,000\,030 \rightarrow$ vedi nota alla linea precedente
$0,006\,2$
$0,016$
$0,046\,2$
$0,092$

BOZZA - DRAFT

$0{,}126\,5$
$0{,}193$
$0,747$
$2,45$
$19,3$
$1\,221,84.$

I 12 valori ordinati sono 4 terne di valori, ed eliminate la prima terna (coi valori più bassi) e l'ultima (coi valori più alti), i 6 valori centrali sono

$0,016$
$0{,}046\,2$
$0,092$
$0{,}126\,5$
$0{,}193$
$0,747$

e la loro media è il valore cercato, la media interquartile, in secondi:

$$\text{IQM} = \frac{0,016 + 0{,}046\,2 + 0,092 + 0{,}126\,5 + 0,193 + 0,747}{6} = \frac{1,220\,7}{6} =$$

$$\boxed{\approx 0,203 \text{ s}}$$

IQM (acronimo di InterQuartile Mean) è un simbolo classicamente usato per la media interquartile; anche iqm e x_{IQM}.

BOZZA - DRAFT

18.9 Inesistenza di una media migliore di tutte[42]

Non esiste una media che sia la migliore di tutte in ogni caso.

Se veramente siamo interessati a produrre

(1) un valore in qualche modo "medio" dei dati

(2) che prescinda dai casi/valori *eccezionali*

allora

- mediana e
- media interquartile

– sono decisamente valide

— ma solo la mediana è *sempre* facilmente calcolabile

∗ invece la media aritmetica richiede di escludere gli outlier:

◇ se i dati sono pochi si può fare a mano ma lasciandoci dubbi

◇ se i dati sono moltissimi non si può proprio fare a mano

però

> si può fare informaticamente con varie regole che definiscono gli outlier, ma in diversi modi a seconda degli Autori.

Avremo sempre bisogno di una qualche media, mica possiamo fermarci a dare liste di migliaia o milioni o miliardi di numeri: è necessario quel passo avanti che riassume i dati in 1 numero medio.

BOZZA - DRAFT

18.10 Disilludersi sulle medie anche senza outlier[46]

Non ci si illuda: non esiste alcun tipo di statistica di sintesi che riassuma bene con 1 solo numero ogni dataset, neppure in assenza di outlier. Spesso la mediana rappresenta bene il soggetto "tipico", ma non sempre: in un cortile con 6 cani e 6 galline, la media e la media interquartile e la mediana del numero di zampe per animale è 3. Che non rappresenta alcun caso tipico. (E la media geometrica è ≈ 2.83, di male in peggio). Nemmeno l'*unimodalità* è una garanzia per la mediana, pur in generale così valida: con 1 grillo, 5 cani e 6 galline

$$6,\ 4,\ 4,\ 4,\ 4,\ 4,\ 2,\ 2,\ 2,\ 2,\ 2,\ 2 \quad (moda = 2)$$

la mediana è ancora 3.

Così, il numero medio di tumori alla prostata e all'utero, di I, II e III grado, in un determinato gruppo di soggetti non ha nessun valore statistico: bisognerà fare statistiche separate per sesso.

BOZZA - DRAFT

18.11 Stratificazione per età in Medicina e Farmacia[50]

Nelle statistiche di Medicina e Farmacia, i dati relativi all'essere umano talvolta vengono variamente stratificati per età e tipicamente sono presentati in:

0-4
5-9
10-14
...
oppure a decenni
10-19
20-29
...
o anche, a ventenni,
20-39
40-59
60-79
...

Si considerano gli anni compiuti, e allora chi avesse 19 anni e 11 mesi sta nella classe 10-19.

BOZZA - DRAFT

18.12 Sul trasformare dati ordinali in dati numerici[54]

La trasformazione di dati ordinali in numerici introduce elementi di arbitrarietà. Vediamo un esempio.

Supponiamo che un'organizzazione raccolga dati sulla soddisfazione degli utenti sulle prestazioni di un suo organo: per esempio un'università, sull'insegnamento di un suo docente – caso reale in Italia. L'utente può scegliere, per valutare il suo grado di soddisfazione, fra

no
più no che sì
più sì che no
sì

Un'università trasforma i valori ordinali in numerici con lo schema

2
5
7

BOZZA - DRAFT

10

e un'altra con quest'altro, altrettanto legittimo:

1

2

3

4

Consideriamo i giudizi sui famosi Professori Pinco e Pallino, dati dai loro 2 studenti per ciascuno:

Pinco: più no che sì, sì: 5+10=15 oppure 2+4=6

Pallino: più sì che no, più sì che no: 7+7=14 oppure 3+3=6.

Col primo schema Pinco figura meglio di Pallino, 15 a 14, e col secondo sono pari, 6 a 6.

BOZZA - DRAFT

18.13 Complementi – Illusioni percettive con le medie

In https://www.epicentro.iss.it/coronavirus/bollettino/Report-COVID-2019_4_ottobre.pdf leggiamo

> Dati al 7 settembre 2020 (...) L'età media dei pazienti deceduti e positivi a SARSCoV-2 è 80 anni

e coi dati al 5 ottobre 2021 è ancora 80 anni. Eppure dall'aprile 2021 al settembre 2021 l'età media dei morti è molto più bassa, come si vede in Figura 3 di https://www.epicentro.iss.it/coronavirus/bollettino/Report-COVID-2019_5_ottobre_2021.pdf. Addirittura, dal grafico vediamo che nel giugno 2020 l'età media era circa 84 anni e nel giugno 2021 (poco meno di) 76. Niente di strano: i morti del periodo estivo sono pochi nel 2020 e molti di più ma comunque pochi nel 2021, la media la fanno principalmente i morti di marzo-aprile 2020 e poi novembre 2020-maggio 2021.

BOZZA - DRAFT

ESERCIZI SULLA LEZIONE 18[58]

ESERCIZIO$_{\mu 2018}$

∗ Considerato il seguente dataset

$$19.68\ 19.20\ 19.63\ 18.94\ 18.81\ 18.10\ 18.63\ 18.85\ 0.01\ 19.51\ 19.54$$

che possiamo supporre misurazioni di parametri corporei, si determini la mediana dopo avere eliminato un outlier.

SVOLGIMENTO

Chiaramente 0.01 è l'outlier preannunciato. (Potrebbe ragionevolmente provenire da un momentaneo malfunzionamento di una macchina che ha prodotto i dati).
I 10 dati rimanenti riordinati in modo crescente sono

$$18.10\ 18.63\ 18.81\ 18.85\ 18.94\ 19.20\ 19.51\ 19.54\ 19.63\ 19.68.$$

I 2 centrali ovvero mediani sono il 5° e il 6°, eliminando 4 da ogni parte:

$$18.94 \quad 19.20$$

la cui media aritmetica è il risultato cercato:

$$\boxed{19.07}$$

BOZZA - DRAFT

(Se non si eliminasse l'outlier – talvolta non è poi così evidente che qualche valore vada scartato – la mediana sarebbe alquanto simile, 18.94, mentre nei 2 casi le medie aritmetiche sono molto più diverse fra loro, rispettivamente 19.089 e 17.3545; la mediana ha il pregio di risentire poco degli outlier, il che in casi complessi con migliaia o milioni di dati, e magari nessuna certezza su quanti e quali sarebbero da considerare outlier, è molto significativo).

ESERCIZIO$_{\mu 2019}$

≈ Si supponga di avere questi dati di un ospedale in anni successivi, relativi ai consumi di un certo farmaco:

1907 4257
1908 3956
1909 3936
1910 4183
1911 4114
1912 4525
1913 4188
1914 4111
1915 4404
1916 4180
1917 0
1918 0

BOZZA - DRAFT

1919 4361
1920 4035

Dopo aver eliminato gli outlier, determinare la media interquartile dei dati sul consumo del farmaco. (Si approssimi all'intero più vicino).

SVOLGIMENTO

Gli outlier sono i due 0 (verosimilmente dovuti alla guerra).

Ordiniamo il dataset rimanente, che ha 12 elementi:
3936, 3956, 4035, 4111, 4114, 4180, 4183, 4188, 4257, 4361, 4404, 4525
Eliminiamo i primi 3 e gli ultimi 3 valori

$$\Big[3936, 3956, 4035, \Big]4111, 4114, 4180, 4183, 4188, 4257\Big[, 4361, 4404, 4525\Big]$$

e la media dei 6 rimanenti è

$$\approx 4172.17$$

e approssimiamo come richiesto:

$$\boxed{4172}$$

Nota. Mostriamo la straordinaria insensibilità della media interquartile agli outlier rispetto alla media aritmetica. I 12 valori non nulli provenivano da una variabile

BOZZA - DRAFT

aleatoria più o meno normale, e poi erano stati aggiunti 2 zeri in corrispondenza ad anni di guerra. Ora, in questo caso è ben evidente che si tratta di outlier per una serie di motivi:

molto diversi dagli altri valori
valori esattamente 0
corrispondenza con anni di guerra.

Ma in altri casi non è così evidente quali valori considerare outlier, per eliminarli "a mano". Allora potremmo fare la media aritmetica di tutti i valori, che però verrebbe fortemente influenzata dagli outlier. Invece la media interquartile dei 12 valori degli ultimi 12 anni, dal 1909 al 1920, senza escludere gli zeri,

1909 3936
1910 4183
1911 4114
1912 4525
1913 4188
1914 4111
1915 4404
1916 4180
1917 0
1918 0
1919 4361

BOZZA - DRAFT

1920 4035

è 4135, alquanto simile a quella di prima. Gli zeri sono finiti nelle porzioni scartate. Insomma la media interquartile ci mostra di cogliere il *vero valore "medio"* della variabile aleatoria retrostante al fenomeno quando esso avviene *normalmente*, e lo fa in un modo *automatico*, con una formula, che non richiede l'esclusione a mano degli outlier, cosa impossibile se i dati sono milioni.

La media aritmetica 3503 di questi stessi 12 valori invece è molto minore (essenzialmente a causa dei 2 zeri) della media 4188 dei 12 valori non nulli iniziali (e qua si vede la sensibilità della media aritmetica agli outlier).

Sarebbe bello mostrare che similmente avverrebbe considerando tutti i 14 valori iniziali, ma sfortunatamente la definizione di media interquartile per campioni con un numero non quadruplo di elementi non è facilissima per un calcolo a mano.

BOZZA - DRAFT

19 Bar chart, istogrammi, e altri diagrammi

19.1 Diagrammi a torta[06]

Il *diagramma a torta* ha un ovvio significato: gli angoli o equivalentemente le aree sono proporzionali contemporaneamente ai valori assoluti o alle frazioni percentuali da rappresentare. (La proporzionalità agli uni è equivalente alla proporzionalità alle altre). Si veda in questo link-> un diagramma a torta di interesse farmaceutico coi valori assoluti riportati (le percentuali no, sono lasciate all'occhio del lettore che vede le fette della torta; può essere un utile esercizio calcolarne qualcuna).
(Più complessa è la situazione in cui oltre alle percentuali relative a 2 o più casi, il che necessita di 2 diagrammi a torta, si voglia rappresentare anche un altro dato per ciascuno dei casi, e allora si faranno diagrammi a torta di diverse dimensioni).
Gli angoli si misurano col goniometro; ma anche a occhio si può fare qualcosa. Il diffuso software Excel fa i diagrammi a torta.

BOZZA - DRAFT

Funzionano veramente bene solo se i valori da rappresentare sono molto pochi, magari 2 o 3, e specialmente se sono molto diversi fra loro, come 66% e 32% e 2% – ma non di più di 2 ordini di grandezza: valori come 90%, 9.9%, 0.07% e 0.03% (l'ultimo ha solo circa 1 decimo di grado e il penultimo 2 decimi e mezzo) non saranno concretamente apprezzabili su un normale foglio a stampa, o schermo di computer.

Spesso si raggruppano in una classe i valori più piccoli, o più grandi.
Si veda un esempio in

https://www.epicentro.iss.it/coronavirus/sars-cov-2-decessi-italia#2

dove in una stessa classe sono i casi "3 o più".

Si evitino come la peste gli ingannevoli diagrammi a torta tridimensionali in rappresentazione prospettica.

BOZZA - DRAFT

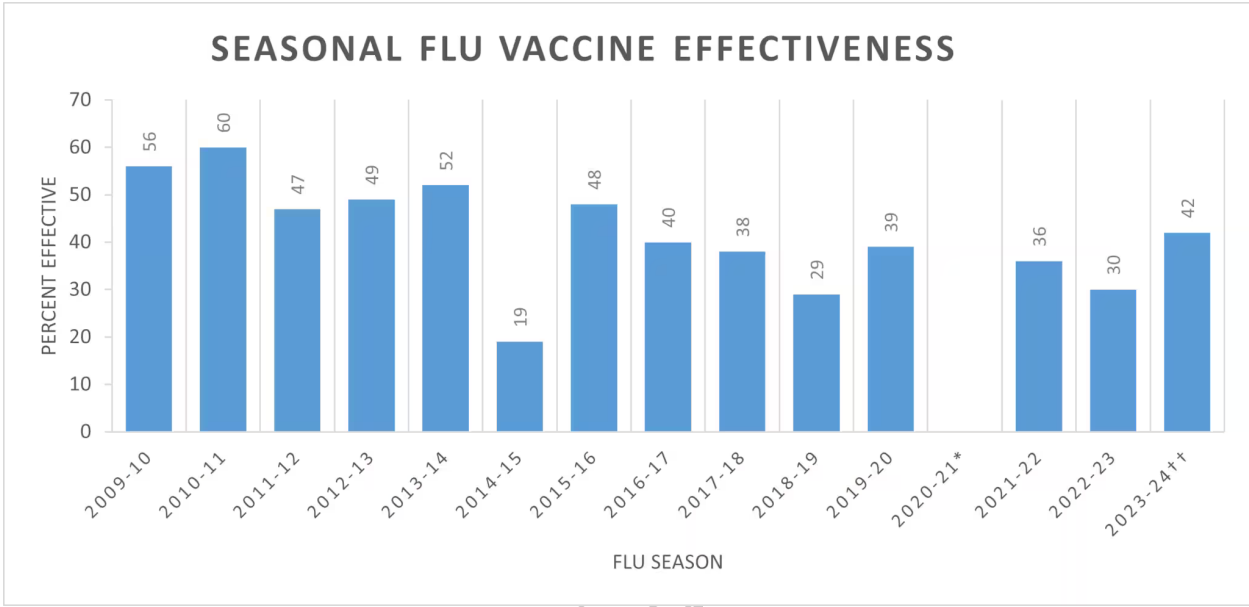
19.2 Catastrofe Statistica: dati mancanti e drop out[10]

Una catastrofe della Statistica è il fenomeno dei dati mancanti.
Si veda un esempio nella prima figura del paragrafo seguente.

Un caso particolare di dati mancanti è quello dei *drop out*, partecipanti che abbandonano un trial clinico prima del suo completamento: sono una catastrofe della Statistica Medica. Ragioni il Lettore: se un cosmetico sperimentale è stato dato a 1000 persone, e dopo un mese rispondono a un questionario 400 persone, e magari risulta che mediamente sono soddisfatte, che valore ha quel risultato statistico? Magari sono alle Seychelles a godersi i risultati, o sono in gravi condizioni e non ci rispondono più al telefono; o sono morti?

BOZZA - DRAFT

19.3 Istogrammi a barre o bar chart, e istogrammi[14]



https://www.cdc.gov/flu-vaccines-work/php/effectiveness-studies/index.html

* 2020-2021 flu vaccine effectiveness was not estimated due to low influenza virus circulation during the 2020-2021 flu season.

I diagrammi a colonne ovvero bar chart ovvero istogrammi a barre, e gli istogrammi – purtroppo con ambiguità nominalistiche nei vari testi – sono diagrammi per la visualizzazione di dati. Nei bar chart l'altezza di ogni colonna – o la lunghezza se disposta orizzontalmente – rappresenta un valore, negli istogrammi l'area rappresenta un valore:

bar chart – altezza ovvero lunghezza – valore

istogramma – area – valore

Spesso in italiano chiamano istogrammi quelli che in questo testo più precisamente chiamiamo bar chart.

Mostriamo un istogramma, nel senso di questo testo (area=valore), e poi ci occuperemo maggiormente di bar chart: LINK->
Si noti che la grandezza sull'asse delle ascisse è continua e non potrebbe essere che così.

BOZZA DRAFT

Torniamo dunque ai bar chart.

Esempio. Dati rilevati a Chimica e Tecnologie Farmaceutiche di Trieste il 20/11/2023, sul numero di gatti e di cani che gli studenti (presenti in aula) hanno:

```
G: 3 2 0 0 2 0 0 0 0 1 0 0 1 0 0 0 0 1 1 1 0 0 0 0 0 0 1 0 0 3 1
C: 0 0 2 0 1 2 0 1 0 0 0 0 0 1 2 0 1 0 1 0 2 0 0 0 1 2 1 2 2 2 1
```

Istogramma a barre delle frequenze per il dataset dei gatti di CTF:

```
  0 gatti:  studenti 20 | ....................
  1 gatto:  studenti 7 | .......
  2 gatti:  studenti 2 | ..
  3 gatti:  studenti 2 | ..
```

BOZZA - DRAFT

19.4 Frequenze assolute, relative, percentuali, cumulate[18]

Dato un dataset di n valori, per ogni valore v che(1) compare:

- la frequenza assoluta f è il numero di volte in cui compare v;
- la frequenza relativa è $\frac{f}{n}$, un numero fra 0 e 1;
- la frequenza percentuale è $\frac{f}{n} \cdot 100\,\%$;

Le frequenze percentuali si possono usare per confrontare due gruppi anche di diversa numerosità, per esempio Farmacia e CTF.

Valgono ovvie(2) relazioni.

BOZZA - DRAFT

[1] C'è poi la frequenza cumulata di v è la somma delle frequenze di tutti i valori *fino a* v compreso (assoluta, relativa o percentuale). La frequenza cumulata ha senso solo per dati almeno ordinali: per dati nominali come pari/dispari non c'è un "fino a".

[2] Essendo k il numero di valori diversi che compaiono nel dataset:

$$f_1 + f_2 + ... + f_k = n \qquad \frac{f_1}{n} + \frac{f_2}{n} + ... + \frac{f_k}{n} = 1$$

Esempio. I gatti degli $n = 31$ studenti di CTF:

gatti	ass.	relativa	percentuale	cumulata ass.	cumulata perc.
0	20	$20/31 \approx 0.645$	64.5%	20	64.5%
1	7	$7/31 \approx 0.226$	22.6%	27	87.1%
2	2	$2/31 \approx 0.065$	6.5%	29	93.5%
3	2	$2/31 \approx 0.065$	6.5%	31	100%
totale	31	1	100%		

Le percentuali arrotondate sommano a 100.1% e non a 100%: è l'effetto degli arrotondamenti, normalissimo, e capita spesso di trovarlo anche nelle statistiche ufficiali. Per questo l'ultima cumulata si scrive 100% e non si ottiene sommando le percentuali arrotondate.

La cumulata si legge così:

il 64.5% *degli studenti ha 0 gatti*;
l'87.1% *degli studenti ha al più 1 gatto*;
il 93.5% *degli studenti ha al più 2 gatti*;
il 100% *degli studenti ha al più 3 gatti*.

BOZZA - DRAFT

ESERCIZIO. Con i dati sui cani degli stessi studenti di CTF (`C: 0 0 2 0 1 2 0 1 0 0 0 0 0 1 2 0 1 0 1 0 2 0 0 0 1 2 1 2 2 2 1`) si costruisca la tabella delle frequenze assolute, relative, percentuali e cumulate, si disegni il bar chart, e si calcoli il numero medio di cani.

SVOLGIMENTO

Contando: 0 cani: 15 studenti; 1 cane: 8; 2 cani: 8. Totale 31.

cani	ass.	percentuale	cumulata ass.	cumulata perc.
0	15	48.4%	15	48.4%
1	8	25.8%	23	74.2%
2	8	25.8%	31	100%

```
0 cani:  studenti 15 | ...............
1 cane:  studenti 08 | ........
2 cani:  studenti 08 | ........
```

Numero medio di cani:

$$\frac{0 \cdot 15 + 1 \cdot 8 + 2 \cdot 8}{31} = \frac{24}{31} \approx 0.77.$$

Confrontando con i gatti: è senza cani il 48.4% degli studenti, senza gatti il 64.5%; e infatti i cani in media sono di più (0.77 contro 0.55).

BOZZA - DRAFT

19.5 Avvertenze sui bar chart[22]

Si noti che è possibile che una colonna rappresenti, per così dire, un valore medio di dati già considerati in altre colonne, per esempio in questo LINK->

Le varie barre possono avere colori che le associano fra loro: LINK->

Ma si faccia attenzione nell'interpretare le statistiche perché molto dipende da come si raggruppano i dati. Nel linkato diagramma sulle cause di morte il cancro compare poco perché è stato disaggregato in vari tipi di cancro. Considerandoli tutti, il cancro è la seconda causa di morte, dopo le malattie cardiocircolatorie. Addirittura qua, disaggregando i tumori in tanti tipi diversi, si fa apparire il covid come prima causa di morte in Italia, datato 3 maggio 2020: LINK->

BOZZA - DRAFT

If you torture the data long enough, it will confess to anything.

(In questo LINK-> è in un articolo su PubMed).

Si faccia attenzione nell'interpretare i bar chart in scala logaritmica perché le altezze rappresentano sì i valori ma non in modo proporzionale ad essi: LINK->

I bar chart possono rappresentare anche valori negativi: LINK->

Naturalmente si faccia attenzione a distinguere i valori assoluti dai valori relativi; per esempio di vari stati o regioni si possono considerare i morti di covid come numero assoluto, o come morti per milione o centomila abitanti. Vediamo un esempio in cui i valori sono già opportunamente scalati sul migliaio di abitanti in questi spettacolari diagrammi a bolle interattivi: LINK->

BOZZA - DRAFT

Si vedano vari tipi di bar chart di interesse specificamente farmaceutico in questo LINK->
Quella sorta di pelucchi che si vedono su certi bar chart rappresentano le deviazioni standard, concetto che vedremo, e che in sostanza rappresentano la variabilità della grandezza rappresentata, all'interno del campione da cui proviene.

BOZZA - DRAFT

19.6 Eventuali asimmetrie nei dataset: skewness[26]

Una distribuzione coi dati più "addensati" verso i valori bassi o quelli alti si dice **skewed**, e di formule e altro non ci occupiamo.

Per il Lettore interessato. In particolare una distribuzione coi dati più "addensati" verso i valori bassi che quelli alti si dice *right skewed*, e si intuisce com'è una distribuzione *left skewed*. (Queste *non* sono definizioni rigorose[3] ma permettono di decidere nella generalità dei casi non particolarmente "capricciosi").
Si provi a dire se è right o left skewed un diagramma a colonne coi dati

12.2 20%
12.4 30%
12.6 25%
12.8 12.5%
13.0 5%
13.2 4%
13.4 2.5%

13.6 1%

BOZZA - DRAFT

[3] Il lettore interessato cercherà sulla rete la *formula*, piuttosto complessa, che quantifica la skewness.

19.7 Funzioni a campana varie[30]

Facendo di istogrammi a barre di dati presi dalla realtà si troverà che spesso le barre si dispongono a formare più o meno una campana.

Praticamente dietro quasi ogni fenomeno della realtà sensibile c'è in qualche modo una funzione più o meno a campana. Imparare a riconoscere queste configurazioni a campana aumenta la comprensione della realtà.

BOZZA - DRAFT

La campana può avere 2 significati principali: una densità, la classica
"distribuzione più o meno a campana",
coi suoi casi estremi rari e quelli medi più comuni,
oppure,
se in ascissa abbiamo il tempo, talvolta rappresenta (quantitativamente) la classica
"parabola della vita",
valida anche per una comunità microbica, per la potenza dell'Impero Romano, per la concentrazione di un farmaco immesso nel sangue, il numero di malati in un'epidemia, o quant'altro: sorgere, ascendere, declinare, finire.

ROZZA - DRAFT

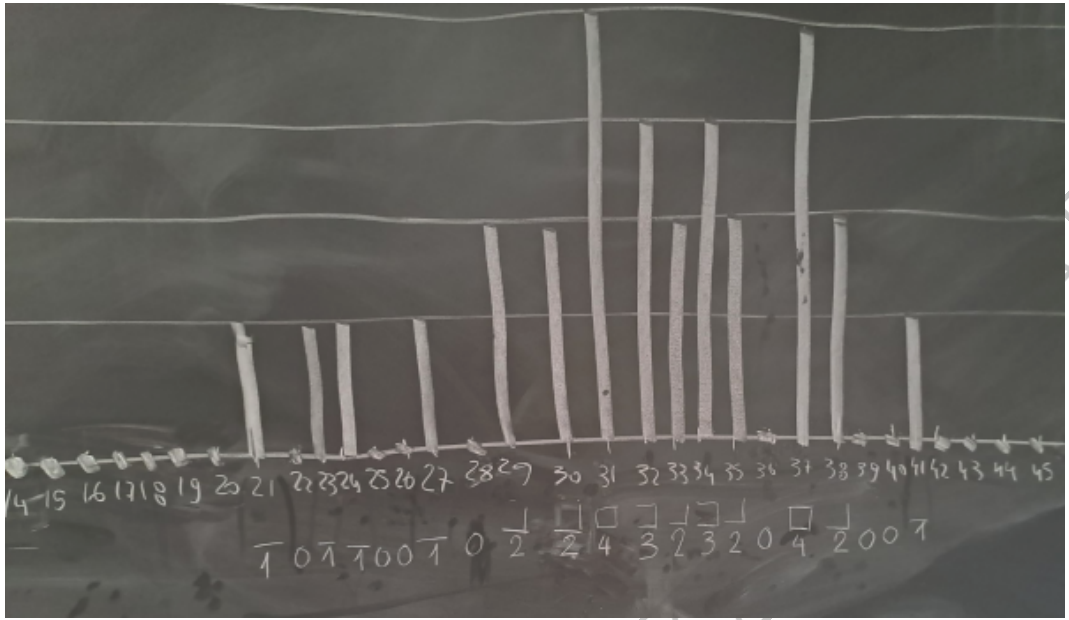


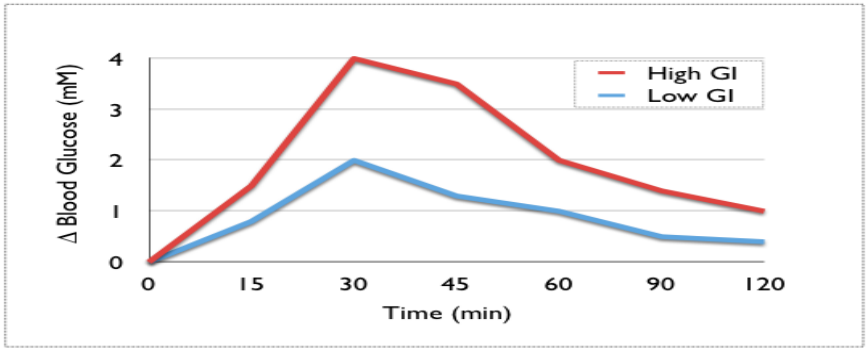
Figura 2: Qual era l'età della madre al momento della tua nascita? Anni compiuti, numero intero. Rilevazione in aula (2023) con questionario anonimo. Anni madre – frequenza: 21 1, 22 0, 23 1, 24 1, 25 0, 26 0, 27 1, 28 0, 29 2, 30 2, 31 4, 32 3, 33 2, 34 3, 35 2, 36 0, 37 4, 38 2, 39 0, 40 0, 41 1.

Qua stiamo parlando di curve "più o meno a campana" mentre con "curva a campana" *di solito* si intende proprio la *campana gaussiana*,

grafico della *densità normale standard* $\frac{1}{\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{t^2}{2}}$, la forma più "pura" di curva a campana del primo tipo.

Curve più o meno a campana del primo tipo, cioè densità (della Microbiologia), sono in questo link->.

Curve più o meno a campana del secondo tipo, cioè evoluzioni di una quantità nel tempo, sono in questa figura.[4]



Altre, relative ad epidemie di influenza, sono[5] queste: LINK->

[4]Immagine di pubblico dominio tratta da Wikimedia Commons.

[5]Tratte da Mid-season real-time estimates of seasonal influenza vaccine effectiveness in persons 65 years and older in register-based surveillance, Stockholm County, Sweden, and Finland, January 2017. Euro

Un'altra (della Microbiologia) è in questo link-> un'altra (della Farmacocinetica) in questo link-> e un'altra a pagina 44 di questo link->.

Leggiamo nel Fedone, XXXIX (IV secolo a.C.), di Platone, citato in Wikipedia, l'enciclopedia libera, alla voce *Variabile casuale*:

> Credi forse che sia tanto facile trovare un uomo o un cane o un altro essere qualunque molto grande o molto piccolo o, che so io, uno molto veloce o molto lento o molto brutto o molto bello o tutto bianco o tutto nero? Non ti sei mai accorto che in tutte le cose gli estremi sono rari mentre gli aspetti intermedi sono frequenti, anzi numerosi?

Nota. Naturalmente nelle Scienze Applicate ricorrono anche grafici ben lontani dall'avere una forma a campana, pur così ubiqua. Già abbiamo visto sigmoidi e sinusoidi... Ma la realtà sensibile è ancora più ricca, per esempio si veda il grafico a questo LINK->.

BOZZA - DRAFT

Surveill. 2017 Feb 23;22(8). pii: 30469. doi: 10.2807/1560-7917.ES.2017.22.8.30469. By Hergens MP et al. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5356437/. "This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence, and indicate if changes were made." LINK->

ESERCIZI SULLA LEZIONE 19[34]

BOZZA - DRAFT

19.7.1 Esercizio risolto a – Bar chart[34]

$_{\mu}$ *

Consideriamo il modello di un fenomeno, e per fissare le idee diciamo che è un'epidemia. (Anche se quello che considereremo non è il miglior modello proprio per le epidemie). Il numero di morti dell'epidemia sia modellizzato da

$$x_n = \left\lfloor \frac{20\, n^2}{n\,!} \right\rfloor \qquad n = 1, 2, ...$$

essendo n il numero di giorni dall'inizio ed indicando il simbolo $\lfloor ... \rfloor$ la parte intera (quella che "toglie i decimali" ai numeri positivi, per esempio $\lfloor 3.14 \rfloor = 3$). Fare un istogramma a barre fino all'ultimo giorno con morti (o per meglio dire fino al giorno precedente del primo

giorno con 0 morti).

SVOLGIMENTO

Usiamo il punto decimale, come già nel testo del quesito (3.14).

$$x_1 = \left\lfloor \frac{20 \cdot 1 \cdot 1}{1} \right\rfloor = 20$$

$$x_2 = \left\lfloor \frac{20 \cdot 2 \cdot 2}{1 \cdot 2} \right\rfloor = 40$$

$$x_3 = \left\lfloor \frac{20 \cdot 3 \cdot 3}{1 \cdot 2 \cdot 3} \right\rfloor = 30$$

$$x_4 = \left\lfloor \frac{20 \cdot 4 \cdot 4}{1 \cdot 2 \cdot 3 \cdot 4} \right\rfloor = \left\lfloor \frac{40}{3} \right\rfloor = \lfloor 13.33... \rfloor = 13$$

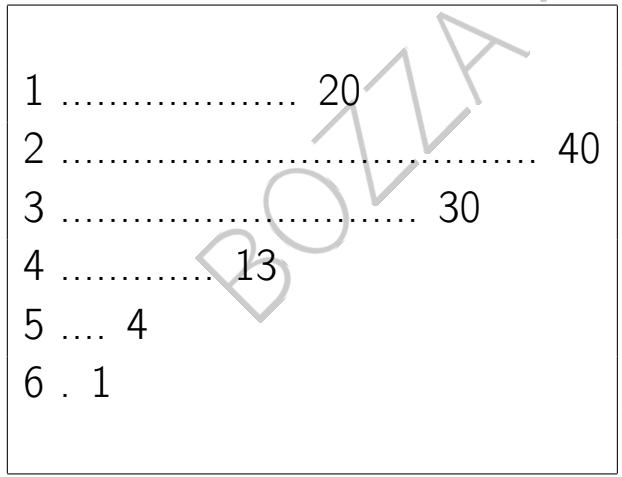
$$x_5 = \left\lfloor \frac{20 \cdot 5 \cdot 5}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} \right\rfloor = \left\lfloor \frac{100}{24} \right\rfloor = \lfloor 4.16... \rfloor = 4$$

$$x_6 = \left\lfloor \frac{20 \cdot 6 \cdot 6}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6} \right\rfloor = 1$$

BOZZA - DRAFT

$$x_7 = \left\lfloor \frac{20 \cdot 7 \cdot 7}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7} \right\rfloor = \left\lfloor \frac{140}{720} \right\rfloor = \lfloor 0.19... \rfloor = 0$$

e abbiamo raggiunto lo 0 previsto. Il numero di morti x_n calcolato dal modello poi non si risolleva da 0 com'è dimostrabile con passaggi facili però laboriosi, ma questa osservazione non ci serve perché nel testo è detto di fermarsi al giorno precedente del primo giorno con 0 morti, che è il giorno numero 7, come abbiamo trovato. Allora l'epidemia dura 6 giorni, e possiamo produrre l'istogramma a barre.



BOZZA - DRAFT

(Che molto meglio sarà rappresentare con le aste disposte verticalmente).

Nota. Ecco una soluzione online su WolframAlpha:

LINK -> (con l'istruzione `DiscretePlot`)

Rispetto a quello dato nella risoluzione dell'esercizio, questo risultato di WolframAlpha

- ha il pregio della precisione
- ha il pregio delle barre disposte verticalmente
- ha il difetto che non scrive sulle barre i 6 valori effettivi.

Ecco un'altra soluzione online su WolframAlpha:

LINK -> (con le istruzioni `BarChart` e `Table`)

Rispetto a quello dato nella risoluzione dell'esercizio, questo risultato di WolframAlpha

- di nuovo, ha il pregio della precisione
- di nuovo, ha il pregio delle barre disposte verticalmente
- ha il pregio che le barre hanno forma di rettangoli e non segmenti
- di nuovo, ha il difetto che non scrive sulle barre i 6 valori effettivi

BOZZA - DRAFT

– ha il difetto che è piuttosto complessa da trovare, richiedendo l'istruzione `Table`

19.7.2 Esercizio risolto b – Istogramma[38]

μ2018
In relazione all'elemento sodio (Na), definite le cellule
ipo-sodiche: con contenuto di sodio fra 0 e $< 1\,000$ unità
normo-sodiche: con contenuto di sodio fra $1\,000$ e $< 10\,000$ unità
iper-sodiche: con contenuto di sodio fra $10\,000$ e $< 20\,000$ unità
(unità che non specifichiamo, essendo ininfluente) un macchinario esamina $52\,000$ cellule (tutti numeri arrotondati per semplicità) trovandovi
$6\,000$ ipo-sodiche, $36\,000$ normo-sodiche, $10\,000$ iper-sodiche.
Rappresentare la situazione con un istogramma. (Non è il bar chart).

BOZZA - DRAFT

SVOLGIMENTO

Nell'istogramma – che purtroppo alcuni Autori scambiano col bar chart ma qua veniamo avvertiti della differenza – le grandezze sono propor-

zionali alle aree dei rettangoli (invece nei bar chart alle lunghezze delle aste, eventualmente strisce rettangolari).

Avremo quindi 3 rettangoli, con basi che si estendono su un asse orizzontale

da 0 a 1000 lunga $1\,000$

da $1\,000$ a $10\,000$ lunga $9\,000$

da $10\,000$ a $20\,000$ lunga $10\,000$.

(Non vi è alcuna questione problematica fra valori compresi o esclusi).

Le altezze le otteniamo dividendo le aree

$6\,000$, $36\,000$, $10\,000$ (o numeri ad essi rispettivamente proporzionali)

per le lunghezze delle basi corrispondenti, ottenendo:

6 unità

4 unità

1 unità

Come unità si può prendere per esempio 1 centimetro, o su carta a quadretti 1 o 2 o 3 quadretti, o anche altre unità.

BOZZA - DRAFT

Ecco l'istogramma, coi 3 rettangoli ed una legenda:

```
A----------
A---------- A=ipo-sodiche, 6 000
ABBBBBBBBB----- B=normo-sodiche, 36 000
ABBBBBBBBB----- C=iper-sodiche, 10 000
ABBBBBBBBB-----
ABBBBBBBBBCCCCCCCCCC
!!----!-----!
0 1000       10000            20000
```

ESERCIZIO.[42] Considerato il dataset

$$2, 3, 1, 4, 1, 5, 3, 4, 7, 1, 2, 4, 5, 9, 4, 8$$

se ne rappresenti il bar chart. Poi si rappresenti l'istogramma con intervalli $[0, 2.5[$, $[2.5, 5[$, $[5, 7.5[$, $[7.5, 10[$, e poi con intervalli $[0, 2.5[$, $[2.5, 5[$, $[5, 10[$.

BOZZA - DRAFT

ESERCIZIO$_{\mu 2019}$ [46]

≈ In un articolo scientifico(6) della rivista Neuropsychopharmacology leggiamo

> Patients were treated two to three times per week with high-frequency MST (i.e., 100 Hz) (N=24), medium frequency MST (i.e., 60 or 50 Hz) (N=26), or low-frequency MST (i.e., 25 Hz MST) (N=36)

Si disegni un istogramma a barre (bar chart) coi valori N_i denominando N l'asse delle ascisse, con barre denominate (lo si scriva entro esse)

high-frequency MST

medium frequency MST

low-frequency MST

e all'estremità di ogni barra si scriva la frequenza relativa, cioè N_i/N_{tot}, espressa percentualmente, con 1 decimale.

BOZZA - DRAFT

[6] *Magnetic seizure therapy (MST) for major depressive disorder*, Neuropsychopharmacology (5 September 2019), Zafiris J. Daskalakis, Julia Dimitrova, Shawn M. McClintock, Yinming Sun, Daphne Voineskos, Tarek K. Rajji, David S. Goldbloom, Albert H. C. Wong, Yuliya Knyahnytska, Benoit H. Mulsant, Jonathan Downar, Paul B. Fitzgerald & Daniel M. Blumberger