Statistica e Probabilità

Esami di Stato e Invalsi

Congresso nazionale Mathesis - Gioia del Colle 2015

di Luciano Corso

E-mail: lcorso@iol.it

Presento alcune osservazioni inerenti allo studio di argomenti di calcolo delle probabilità e di statistica,

e la risoluzione di alcuni
problemi dati agli esami di Stato dei Licei Scientifici e indicati nelle prove INVALSI

Differenza tra popolazione e campione

- Popolazione: Dato un certo carattere statistico (variabile o mutabile che sia), per popolazione si intende la totalità dei dati sperimentabili che ne descrivono il comportamento. All'interno di questo gruppo di dati si esaurisce tutta l'informazione inerente a quel carattere.

- Campione: dato un certo carattere statistico (variabile o mutabile che sia), per campione si intende una raccolta di dati (un gruppo di essi) tratti dalla popolazione. A seconda che si estraggano i dati con o senza reinserimento e del tipo di popolazione sotto osservazione la cardinalità del campione può essere minore, uguale o maggiore dei dati della popolazione.

Osservazioni:

- Una popolazione può essere di numerosità finita o infinita e così il campione. La differenza sta nel fatto che normalmente un campione presenta una distorsione nell'informazione che contiene rispetto alla popolazione. Non c'è invece alcun vincolo riguardo alla numerosità campionaria che non è detto che sia minore di quella della popolazione.
- Facciamo un esempio per chiarire la questione. Da una popolazione costituita da $N = 6$ palline numerate presenti in una scatola e riportanti i numeri $\{1, 2, 3, 4, 5, 6\}$ si estraggono, con esperimenti diversi rispettivamente $n = 2$ palline, $n = 6$ palline, $n = 10$ palline. L'estrazione sia con reinserimento e poi senza. Si vuole eseguire l'esperimento per stimare la media aritmetica della popolazione.

Campioni con n = 2 estratti con prove indipendenti da una popolazione N=6

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

- n = 2 è minore di N = 6. Rispetto alla stima della media della popolazione solo 6 campioni su 36 centrano l'obiettivo. I campioni presentano in genere una distorsione nella ricerca della media aritmetica della popolazione.

Campioni n = 6 ed estrazione in blocco e con reinserimento da una popolazione lunga N=6

- Il campione 123456 è il caso di un campione che corrisponde esattamente alla popolazione e ci sono 6! campioni di questo tipo. In questo caso e solo in questo caso la media del campione corrisponde alla media della popolazione. Ciò significa che si è in presenza di un campione che raccoglie tutta l'informazione della popolazione.

- I campioni sono in tutto 6^6 = 46656; tra queste ce ne sono 6! = 720 che raccolgono totalmente i dati della popolazione. Anche in questo caso i campioni sono generemalmente distorti anche se ce ne sono 6! Che sono uguali alla popolazione. I campioni hanno uguale numerosità della popolazione.

Campioni con n = 10 estratti da una popolazione lunga N=6 con prove indipendenti (reinserimento)

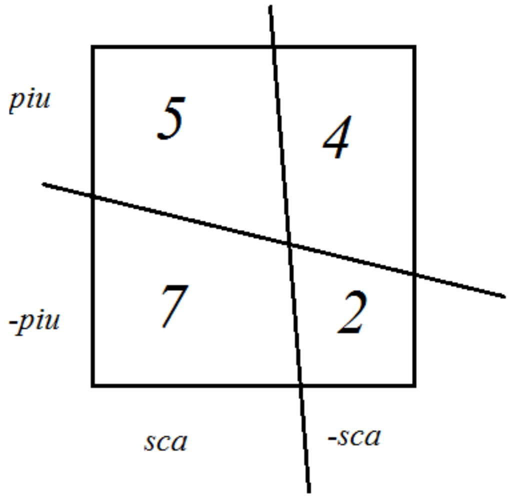
In questo caso si deve formare il campione con estrazioni con reinserimento. I campioni possibili sono 6^10 = 6046176. Uno di questi campioni è 1223455566.

La media di questo campione è m = 3,9 (la media aritmetica della popolazione è μ = 3,5. Si può notare che il campione ha una numerosità maggiore della popolazione.

Problema 1: sullo spazio campionario

In un gruppo di ragazzi è stato osservato che 12 calzavano scarponi; 9 indossavano il piumino; 5 calzavano scarponi e indossavano il piumino; infine, 2 ragazzi non indossavano il piumino né calzavano scarponi. Si domanda: 1) il numero dei ragazzi; 2) quanti calzavano gli scarponi e non avevano il piumino; 3) quanti indossavano il piumino ma non calzavano scarponi.

Risoluzione: La prima cosa da fare in questo tipo di problemi è di rappresentare i possibili eventi in un diagramma di Venn modificato. Lo spazio campionario diventa un quadrato di lato unitario.
Le parti della partizione dello spazio campionario Ω *sono qui sotto rappresentate per costruzione. Lo studente impara che* $|\Omega|=2^2$ *e in generale che* $|\Omega| = 2^n$ *dove n sono i caratteri sotto analisi.*



Problema 2. Esame di Stato sessione ordinaria 2015: Quesito n. 3
Lanciando una moneta sei volte qual è la probabilità che si ottenga testa "al più" due volte? Qual è la probabilità che si ottenga testa "almeno" due volte?

Osservo che l'uso del concetto di variabile aleatoria facilita la ricerca della risposta corretta. In questo caso x = «numero di volte che esce testa in n=6 lanci di moneta ideale».
Risoluzione

$$P\left(x \le 2 \middle| n = 6,\ p = 1/2\right) = \sum_{x=0}^{2} \binom{6}{x} \cdot \left(\frac{1}{2}\right)^{x} \left(1 - \frac{1}{2}\right)^{6-x} = \frac{22}{64}$$

$$P\left(x \ge 2 \middle| n = 6,\ p = 1/2\right) = \sum_{x=2}^{6} \binom{6}{x} \cdot \left(\frac{1}{2}\right)^{x} \left(1 - \frac{1}{2}\right)^{6-x} = \frac{57}{64}$$

Problema 3. Esame di Stato 2013 sessione ordinaria – corso sperimentale: quesito n. 6

- Con le cifre da 1 a 7 è possibile formare 5040 numeri corrispondenti alle permutazioni delle 7 cifre. Ad esempio i numeri 1234567 e 3546712 corrispondono a due di queste permutazioni. Se i 5040 numeri ottenuti dalle permutazioni si dispongono in ordine crescente qual è il numero che occupa la 5036-esima posizione e quale quello che occupa la 1441-esima posizione?

Il problema è di calcolo combinatorio. La risposta stimola la riflessione sull'ordinamento dei numeri, imponendo una netta distinzione tra cifre e numeri.

- Risoluzione:
- Risposta al primo quesito: ordinando i numeri più grandi a partire dal più grade e considerando che sono 3! I modi di formare questi numeri permutando le ultime tre cifre 1,2,3. Il gioco è fatto:
- 7654 321 , 7654 312 , 7654 231 , 7654 213 , 7654 132 , 7654 123 ,
  5040 , 5039 , 5038 , 5037 , 5036 , 5035 , …
- Risposta al secondo quesito: tenendo ferma la prima cifra e permutando le altre 6, a partire dal numero più piccolo si ha:
- 1 : 234567 e sono 6! = 720 i numeri distinti che si possono formare,
  2 : 134567 e sono anche qui 6! = 720 i numeri distinti che si possono formare, e
- quindi in totale 1440 numeri sulle prime due cifre,
  3 : 124567 è il 1441-esimo numero della sequenza non decrescente.

Problema 4. Esame di Stato 2015, quesito n. 7

In un gruppo di 10 persone il 60% ha occhi azzurri. Dal gruppo si selezionano a caso due persone. Quale è la probabilità che nessuna di esse abbia occhi azzurri?

Risoluzione: è diretta, sugli eventi dipendenti

$P(–A_1 \text{ e } –A_2) = (4/10)\,(3/9) = 12 / 90.$

Si può procedere anche con il metodo delle combinazioni semplici:

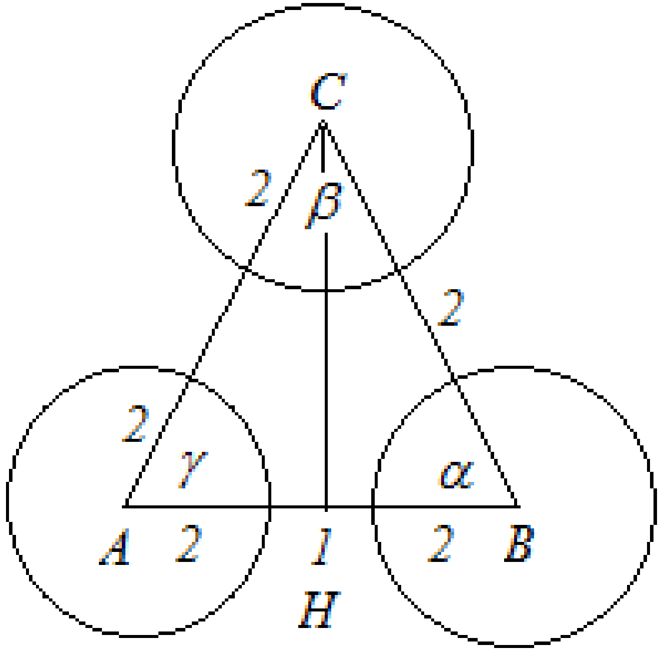
$$\binom{4}{2} \Big/ \binom{10}{2}$$

Problema 5. Esame di Stato 2015, quesito n. 8

I lati di un triangolo misurano, rispettivamente, 6 cm, 6 cm e 5 cm. Preso a caso un punto P all’interno del triangolo, qual è la probabilità che P disti più di 2 cm da tutti e tre i vertici del triangolo?

- Risoluzione:
- La risposta non è immediata. Richiede una conoscenza di trigonometria, geometria, probabilità. Vi sono almeno due percorsi per risolvere il quesito (si veda la proposta di Luigi Verolino su MatMedia). Io seguo questo:

Risoluzione (secondo me; si veda anche MatMadia – L. Verolino):



$$\frac{6}{\sin\alpha} = \frac{5}{\sin\beta}\ ;\ \frac{6}{\sin\alpha} = \frac{5}{\sin(\pi - 2\alpha)}$$

$$\frac{6}{\sin\alpha} = \frac{5}{2\sin\alpha\cdot\cos\alpha}\quad ;\quad \cos\alpha = \frac{5}{12}\quad ;\quad \alpha = arc\cos\frac{5}{12}$$

|CH| = √(62 – 2,52) ≅ 5,45436

Area.tri = 5 · 5,45436 / 2 ≅ 13,6358

$\alpha = \gamma$

$\alpha = 1{,}1410.$

$\pi\, r^2 : 2\,\pi = x : \alpha \quad ; \quad x = \pi \cdot 2^2 \cdot 1{,}1410 / 2\cdot\pi \cong 2{,}2820$

$2\,x = 2 \cdot 2{,}2820 \cong 4{,}5640.$

Calcoliamo l'area del settore circolare corrispondente all'angolo β.

$\pi\, r^2 : 2\,\pi = y : \beta \quad ; \quad y = \pi\cdot 2^2 \cdot 0{,}8595 / 2\,\pi \cong 1{,}719$

Area dei settori circolari = 2x + y = 4,5640 + 1,719 = 6,283

Casi favorevoli: |A|=13,6358 – 6,2830 = 7,3528

Casi Favorevoli |Ω| = 13,6358

Pr(A) = 7,3528 / 13,6358 = 0,5392

Problema 6. Piano Nazionale di Informatica sessione 2014 – esami di Stato – Y557 Quesito n. 9.

Un mazzo di “tarocchi” è costituito da 78 carte: 22 carte figurate, dette “Arcani maggiori”, 14 carte di bastoni, 14 di coppe, 14 di spade e 14 di denari. Estraendo a caso da tale mazzo, l’una dopo l’altra con reinserimento, 4 carte, qual è la probabilità che almeno una di esse sia un “Arcano maggiore”?

Risoluzione: anche in questo caso uso il concetto di variabile aleatoria x = «numero di Arcani in n= 4 prove indipendenti».

$$p\left(x \geq 1 \middle| |A| = 22, |\Omega| = 78, n = 4\right) = \sum_{x=1}^{4} \binom{4}{x} \cdot \left(\frac{22}{78}\right)^{x} \cdot \left(\frac{56}{78}\right)^{4-x} \cong 0,73431$$

Oppure, ancor meglio:

$$p\left(x \geq 1 \middle| |A| = 22, |\Omega| = 78, n = 4\right) = 1 - P\left(x = 0 \middle| \ldots\right) = 1 - \binom{4}{0} \cdot \left(\frac{22}{78}\right)^{0} \cdot \left(\frac{56}{78}\right)^{4} \cong 1 - 0,26569 \cong 0,73431.$$

Problema 7. Piano Nazionale di Informatica sessione 2014 – esami di Stato – Y557 – quesito n. 10

Nel poscritto al suo racconto "Il Mistero di Marie Rogêt", Edgar Allan Poe sostiene che, "avendo un giocatore di dadi fatto doppio sei per due volte consecutive, vi è una ragione sufficiente per scommettere che gli stessi sei non usciranno ad un terzo tentativo". Ha ragione? Si motivi esaurientemente la risposta.

Risoluzione quesito n. 10:

Risoluzione quesito n. 10

No. Ha torto. Una cosa è chiedersi qual è la probabilità che in tre prove di due lanci di dadi l'una, escano tre coppie di sei, altra cosa è calcolare la probabilità che al terzo lancio di una coppia di dadi esca il doppio sei, dato che nelle due prove precedenti sono uscite due coppie di sei. Troviamo:

$P(A_1 \text{ e } A_2 \text{ e } A_3 \mid p = 1/36, n = 3) = (1 / 36)^3$.

$P(A_3 \mid A_1 \text{ e } A_2) = P(A_3) = 1 / 36$.

L'errore fatto da Poe è di non aver considerato che l'evento "esce un doppio sei" non ha memoria. Poe pensa di poter usare, erroneamente, la legge dei grandi numeri che il più delle volte, come in questo caso, è applicata male. La legge è:

$\lim_{n\to+\infty} \Pr(|f_n - p| \geq \varepsilon) = 0$ con $f_n = x_n / n$.

La legge dice che nel lungo andare una frequenza relativa converge alla probabilità dell'evento considerato, ma non dice che ci si può fidare di ciò all'n-esimo lancio.

Problema 8. Piano Nazionale di Informatica sessione 2014 – esami di Stato – Y557 – quesito n. 3

Venti palline sono poste in un’urna. Cinque sono rosse, cinque verdi, cinque gialle e cinque bianche. Dall’urna si estraggono a caso, senza reimbussolamento, tre palline. Si valutino le seguenti probabilità:

- Esattamente una pallina è rossa
- Le tre palline sono di colori differenti.

Risoluzione quesito n.3:

Risoluzione quesito n. 3

Posto x = numero di palline rosse estratte si ottiene:

$$P\left(x = 1 \mid n = 3, N = 20, C = 5\right) = \frac{\binom{C}{x}\binom{N-C}{n-x}}{\binom{N}{n}} = \frac{\binom{5}{1}\binom{20-5}{3-1}}{\binom{20}{3}} = \frac{35}{76}$$

Posto R = rosso, V = verde, G = giallo, B = bianco, l’evento favorevole è:
(RVG ∪ RVB ∪ RGB ∪ VGB∪ *alle loro permutazioni*) ; la sua probabilità è:

$$\Pr(RVG \cup RVB \cup RGB \cup VGB \cup alle\ loro\ \ permutazioni) = 3!\binom{4}{3}\left(\frac{5}{20}\cdot\frac{5}{19}\cdot\frac{5}{18}\right) = \frac{25}{57}$$

Problema 9. Piano Nazionale di Informatica sessione 2014 – esami di Stato – Y557 – Quesito n. 8. La “zara” è un gioco d’azzardo di origine araba che conobbe particolare fortuna in Italia in epoca medievale – ne parla anche Dante nella Divina Commedia – e si giocava con tre dadi. Si confronti la probabilità di ottenere in un lancio la somma 9 con quella di ottenere la somma 10.

Si veda il discorso di G. Galilei su questo problema (esemplare per la chiarezza).

9	126	162	216	261	612	621	225	252	522
	324	342	234	243	423	432	333	441	414
	144	531	513	135	153	315	351		
10	136	163	316	361	613	631	226	262	622
	235	253	325	352	523	532	244	424	442
	334	343	433	415	451	145	154	514	541

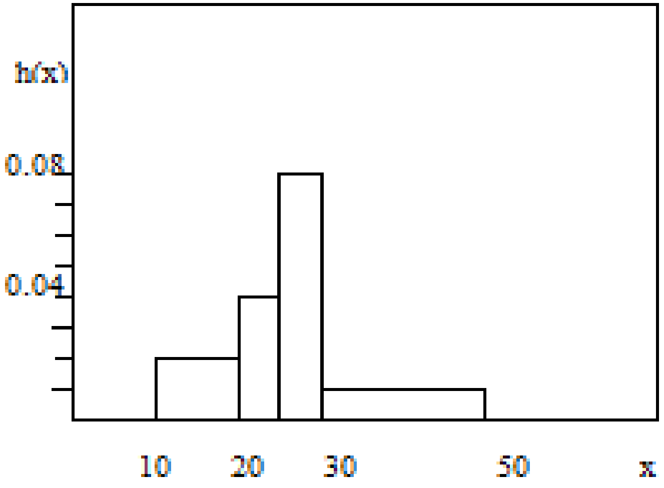
Pr(esce il 9) = 25 / 216 Pr(esce il 10) = 27 / 216

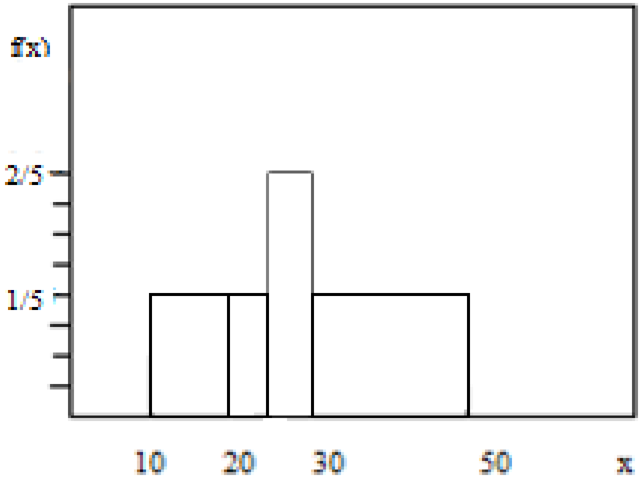
Test ingresso UNIVR – Biotecnologie

Una certa variabile statistica continua X è stata classificata secondo la seguente tabella

Classe	Frequenza Assoluta	Frequenza relativa	Densità di frequenza
[10,20)	10	1 / 5	0,02
[20, 25)	10	1 / 5	0,04
[25, 30)	20	2 / 5	0,08
[30. 50]	10	1 / 5	0,01
	50	1	

Presentare l'istogramma che descrive la distribuzione delle frequenze relative di X.





Il quadro di riferimento Invalsi è buono

[...] Il riferimento alla valutazione per la matematica è quindi costituito dall'asse culturale matematico; in esso si dice che:
la competenza matematica comporta la capacità e la disponibilità a usare modelli matematici di pensiero (dialettico e algoritmico) e di rappresentazione grafica e simbolica (formule, modelli, costrutti, grafici, carte), la capacità di comprendere ed esprimere adeguatamente informazioni qualitative e quantitative, di esplorare situazioni problematiche, di porsi e risolvere problemi, di progettare e costruire modelli di situazioni reali. Finalità dell'asse matematico è l'acquisizione al termine dell'obbligo d'istruzione delle abilità necessarie per applicare i principi e i processi matematici di base nel contesto quotidiano della sfera domestica e sul lavoro, nonché per seguire e vagliare la coerenza logica delle argomentazioni proprie e altrui in molteplici contesti di indagine conoscitiva e di decisione.

Le quattro competenze di base di questo asse culturale sono così enunciate

1) Utilizzare le tecniche e le procedure del calcolo aritmetico e algebrico, rappresentandole anche sotto forma grafica.
2) Confrontare e analizzare figure geometriche, individuando invarianti e relazioni.
3) Individuare le strategie appropriate per la soluzione di problemi
4) Analizzare dati e interpretarli sviluppando deduzioni e ragionamenti sugli stessi anche con l'ausilio di rappresentazioni grafiche, usando consapevolmente gli strumenti di calcolo e le potenzialità offerte da applicazioni specifiche di tipo informatico.

I saperi sono articolati in abilità/capacità e conoscenze, con riferimento al sistema di descrizione previsto per l'adozione del Quadro Europeo dei Titoli e delle Qualifiche (EQF)2 Nel momento in cui una domanda è proposta per il Servizio Nazionale di Valutazione, si cerca di esplicitare in che modo tale domanda possa fornire indicazioni significative riguardo a una di queste competenze di base. Queste informazioni vengono raccolte nelle Guide alla lettura che accompagnano le prove. . Per quanto riguarda il sistema dei licei, gli obiettivi di apprendimento specifici sono contenuti nel complesso dei documenti delle Indicazioni nazionali3 . In essi gli obiettivi di apprendimento sono articolati in quattro ambiti (Aritmetica e algebra; Geometria; Relazioni e funzioni; Dati e previsioni). Per ogni domanda proposta per il Servizio Nazionale di Valutazione viene stabilito un collegamento con uno o più di questi obiettivi di apprendimento e queste informazioni sono riportate nelle Guide alla lettura.

Problema n. 10

Viene richiesto di rappresentare la distribuzione di frequenza e la funzione di ripartizione (distribuzione di frequenza cumulata) della variabile raggruppata in cassi di frequenza della tabella della slide successiva. Ecco ciò che può accadere usando excel.

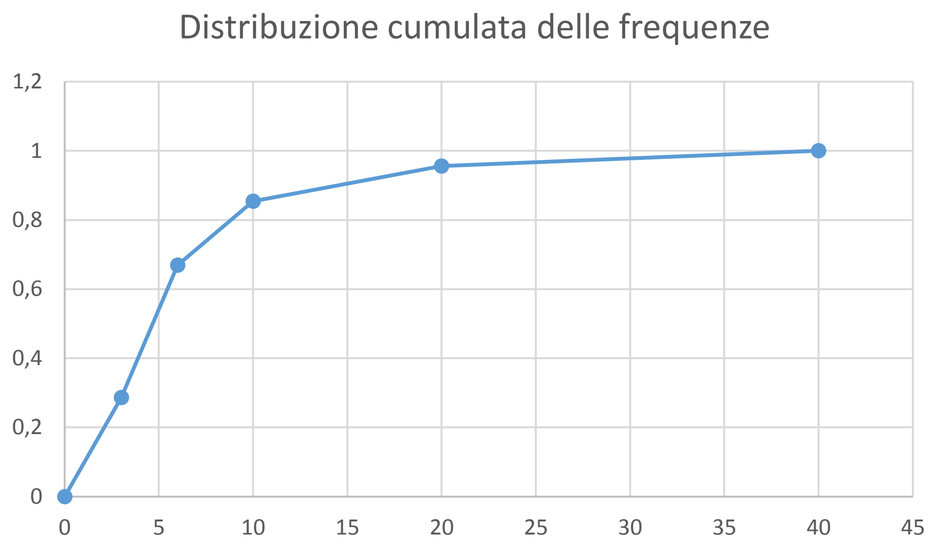
Si critichino i grafici della slide 20 (successiva a questa).

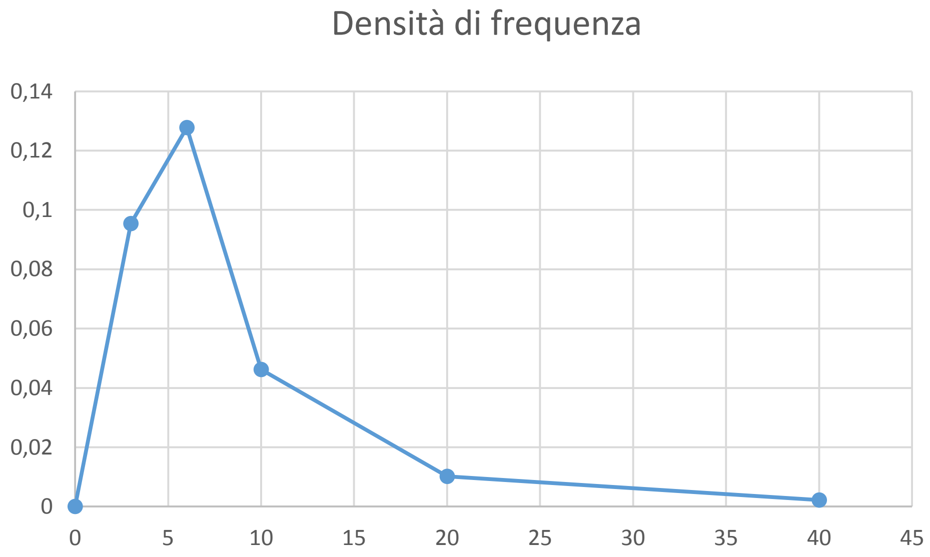
1) Consideriamo il grafico della frequenze cumulate (distribuzione cumulata delle frequenze): il grafico deve dimostrare che è coinvolto tutto l'asse R perché la funzione di ripartizione delle frequenze (che poi è nota anche con il nome di funzione della distribuzione delle frequenze cumulate o semplicemente funzione della distribuzione delle frequenze) è definita su tutto R. Excel dà grafici che sembrano esistere solo in un campo limitato di R. Perciò il grafico si presenta lacunoso.

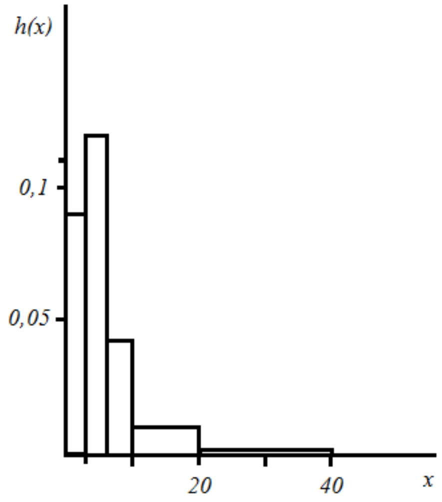
2) Consideriamo il grafico della densità di probabilità fatto con excel. Questo grafico è sbagliato perché all'interno di ogni classe il grafico non si presenta costante. Per esempio, nell'intervallo [0, 3) è crescente, mentre per l'ipotesi di equi distribuzione dei dati all'interno di ogni classe esso dovrebbe essere un segmento costante di retta di ordinata pari alla densità di frequenza della classe. Il grafico in angolo a destra in basso è invece corretto.

SI VEDA LA SLIDE SUCCESSIVA

Classi	n(i)	f(i)	h(i)	xc(i)	F(j)
[0, 3)	65	0,286344	0,095448	1,5	0,286344
[3, 6)	87	0,38326	0,127753	4,5	0,669604
[6, 10)	42	0,185022	0,046256	8	0,854626
[10, 20)	23	0,101322	0,010132	15	0,955947
[20, 40]	10	0,044053	0,002203	30	1
	227	1			







Conclusioni:

- Esami di Stato
- I quesiti sono più complessi e richiedono una visione anche interdisciplinare. Si chiede una preparazione in geometria, calcolo delle probabilità, in trigonometria e in calcolo combinatorio.

- Invalsi
- I quesiti sono rapidi e si richiede padronanza linguistica con un buon vocabolario della disciplina oggetto d'indagine. Si notano più quesiti di statistica rispetto agli esami di stato. I quesiti a risposta multipla sono strutturati in modo da evidenziare le capacità di calcolo e il controllo delle dimensioni di scala.