Istituto Nazionale di Documentazione, Innovazione e Ricerca Educativa
MIUR

I campioni… si contano

di G. Baruzzo, A. D'Arpino, P. Ranzani

Area tematica

Geometria

Autori

G. Baruzzo, A. D'Arpino, P. Ranzani

Ordine di scuola

Secondaria di secondo grado - preferibilmente classe IV

Tempo medio per svolgere il percorso

8-10 ore

Indice

Scheda generale

Nucleo tematico

Dati e previsioni

Autori

G. Baruzzo, A. D'Arpino, P. Ranzani

Tematica affrontata

Costruire lo spazio dei campioni casuali semplici a partire da una popolazione di piccola numerosità. Descrivere la variabile casuale media campionaria, facendo anche uso della simulazione. Sulla base del campione osservato, fornire informazioni sulla media di una popolazione. Elementi di calcolo combinatorio come base per differenti tecniche di campionamento casuale.

Descrizione

I mezzi di comunicazione di massa ci hanno abituati all'uso dei sondaggi i cui risultati dipendono da una indagine campionaria. Il ricorso ad esempi tratti dalla vita quotidiana motiverà gli studenti al tema di questa attività che intende sviluppare tecniche di campionamento probabilistico semplice al fine di evidenziarne le caratteristiche principali e di comprenderne i limiti, con particolare riferimento alla stima della media.

Ordine di scuola

Secondaria di secondo grado: secondo biennio (preferibilmente classe quarta).

Tempo medio per svolgere l'attività in classe

Indicativamente 8/10 ore

Nodi concettuali

- Rudimenti di calcolo combinatorio come requisito al campionamento casuale.
- Campionamento casuale e variabili casuali.
- Ragionamento induttivo e basi concettuali dell'inferenza.
- Campionamento casuale semplice e inferenza induttiva sulla media.

Indicazioni curricolari

Le attività M@t.abel hanno precisi obiettivi di apprendimento che rientrano tra quelli inseriti nelle Indicazioni nazionali attualmente in vigore (D.M. n. 211 del 07/10/2010, Direttiva n. 4 del 16/01/2012, Direttiva n. 5 del 16/01/2012) e nelle Prove INVALSI. All'inizio di ciascuna attività sono riportati, perciò, i relativi riferimenti presenti nelle Indicazioni nazionali e alcuni quesiti delle Prove Invalsi che ripropongono la situazione stimolo dell'attività considerata. Una domanda Invalsi può aiutare a valutare se gli allievi hanno sviluppato, attraverso lo svolgimento dell'attività, la capacità di utilizzare la matematica per rispondere a domande in una situazione specifica. Le domande sono tratte tra quelle presenti nei vari livelli scolastici, in quanto le attività M@t.abel sono pensate in un'ottica di verticalità.

Indicazioni Nazionali per i Licei

Linee generali e competenze

Concetti e metodi che saranno obiettivo dello studio:

- la conoscenza elementare di alcuni sviluppi della matematica moderna, in particolare degli elementi del calcolo delle probabilità e dell'analisi statistica.

Obiettivi specifici di apprendimento

Dati e previsioni

Lo studente, in ambiti via via più complessi, il cui studio sarà sviluppato il più possibile in collegamento con le altre discipline e in cui i dati potranno essere raccolti direttamente dagli studenti, apprenderà a far uso delle distribuzioni doppie condizionate e marginali, dei concetti di deviazione standard, dipendenza, […], e di campione. Studierà […] gli elementi di base del calcolo combinatorio. In relazione con le nuove conoscenze acquisite approfondirà il concetto di modello matematico.

Linee Guida per gli Istituti Professionali e Tecnici

Le **competenze** acquisite consentiranno allo studente di:

- utilizzare il linguaggio e i metodi della matematica per organizzare e valutare adeguatamente informazioni qualitative e quantitative;
- utilizzare i concetti e i modelli delle scienze sperimentali per investigare fenomeni sociali […] per interpretare dati.

Conoscenze

- Campionamento casuale semplice e inferenza induttiva sulla media […].
- Costruire un campione casuale semplice data una popolazione. Costruire stime puntuali ed intervallari per la media […].
- Utilizzare e valutare criticamente informazioni statistiche di diversa origine con particolare riferimento ai giochi di sorte e ai sondaggi. Realizzare ricerche e indagini di comparazione, ottimizzazione, andamento, ecc., collegate alle applicazioni d'indirizzo.

Prove INVALSI

a.s. 2012/2013 - Domanda D28

Scuola secondaria di II grado – Classe II

Il quesito può essere considerato propedeutico alle tematiche svolte nell'unità.

D28. **Un gruppo di biologi, per stimare quante trote ci sono in un lago, ne pesca 200 e, dopo averle marcate, le rigetta nel lago.**
Dopo qualche giorno, utilizzando la stessa rete, vengono pescate 720 trote e solo 12 di esse sono marcate.
In base a queste informazioni, quante trote possiamo pensare che ci siano all'incirca nel lago?

A. ☐ 2000

B. ☐ 9000

C. ☐ 12 000

D. ☐ 144 000

Soluzione INVALSI

D28: C

Commento

Lo scopo della domanda è quello di fare una stima corretta su base campionaria. L'esercizio può essere utile per introdurre i concetti di popolazione, di campione e di stima trattati in questa unità.

Gli studenti possono arrivare alla risposta corretta con la seguente proporzione: 720 : 12 = x: 200.

Da cui $x = \dfrac{720 \cdot 200}{12} = 12000$

Oppure è possibile calcolare direttamente la frazione 720/12 e moltiplicarla per 200.

Analogamente è possibile ricavare il risultato imponendo l'uguaglianza 12/200 = 720/x da cui $x = \dfrac{720 \cdot 200}{12} = 12000$

Si tratta di un problema di proporzionalità che gli studenti sono abituati ad affrontare già dalla scuola secondaria di primo grado. È però probabile che possano incontrare qualche difficoltà nell'accettare la plausibilità dell'ipotesi del modello di proporzionalità diretta o il saper riconoscere in contesti diversi il carattere misurabile di oggetti e fenomeni.

a.s. 2013/2014 - Domanda D12

Scuola secondaria di II grado – Classe II

Il quesito può essere considerato propedeutico alle tematiche svolte nell'unità.

D12. È stato effettuato un sondaggio su un campione di 1 500 donne di età compresa tra i 25 e i 55 anni per conoscere la loro opinione su una rivista mensile dedicata alla salute. Si sono ottenuti i seguenti risultati:

	Occupate	Disoccupate
Giudizio positivo	450	276
Giudizio negativo	367	407

a. Quante sono le donne che hanno espresso un giudizio positivo?

Risposta: ....

b. Quante sono le donne disoccupate intervistate?

Risposta: ....

c. Scegliendo a caso una delle donne intervistate, qual è la probabilità che abbia espresso un giudizio negativo?

Risposta: ....

d. Scegliendo a caso una delle donne intervistate tra quelle che hanno espresso un giudizio positivo, qual è la probabilità che sia una donna occupata?

Risposta: ....

Soluzione INVALSI

D12_a: 726 D12_b: 683

D12_c: 774/1500 = 129/250 = 0,516 = 51,6%

D12_d: 450/726 = 75/121 ≈ 0,6198 cioè circa il 61.98%

Commento

Lo scopo della domanda quello di saper leggere ed utilizzare dati statistici ricavati da una tabella a doppia entrata.

Per rispondere correttamente alle domande **a.** e **b.** è sufficiente saper leggere una tabella a doppia entrata ed eseguire correttamente addizioni tra numeri interi.

Per rispondere correttamente alla domanda **c.** è sufficiente pensare alla probabilità come rapporto tra le donne che hanno espresso un giudizio negativo e il totale delle donne intervistate.

Anche per rispondere alla domanda **d.** è sufficiente sapere calcolare un rapporto, ma in questo caso è necessario riconoscere che l'insieme universo non è più costituito da tutte le donne intervistate, ma solo dalle donne intervistate che hanno espresso un giudizio positivo. Da ciò segue il significato di evento condizionato, probabilità condizionata e loro utilizzo. Il quesito può essere utilizzato dal docente anche per introdurre il concetto di campione e di lavorare su dati campionari.

a.s. 2013/2014 - Domanda D20

Scuola secondaria di II grado – Classe II

Il quesito può essere considerato propedeutico alle tematiche svolte nell'unità.

D20. **Da un controllo di qualità è emerso che una macchina ha prodotto 14 pezzi difettosi su una produzione di 1 200 pezzi. Che stima è ragionevole fare del numero di pezzi difettosi su una produzione di 2 150 pezzi?**

Scrivi i calcoli che hai fatto per trovare la risposta e poi riporta il risultato approssimandolo all'unità.

....

....

....

Risultato (approssimato all'unità): ....

Soluzione INVALSI

D20: 25 (accettabile anche 26)

Commento

Scopo della domanda è quello di stimare una frequenza assoluta a partire da una probabilità frequentista e quindi è anche è possibile introdurre i concetti di popolazione, di campione e di stima trattati in questa unità.

Per rispondere correttamente è sufficiente calcolare la frequenza relativa del numero di pezzi difettosi, cioè 14/1200. Tale frequenza può essere considerata come una stima della probabilità di produrre un pezzo difettoso. Su questa base è ragionevole attendersi che su 2150 pezzi prodotti ve ne siano cioè 25 pezzi approssimando all'intero più vicino o troncando all'unità, oppure 26, se si approssima per eccesso all'intero (commettendo in tal caso un errore di approssimazione maggiore di quello commesso quando si è approssimato il risultato con 25)..

Introduzione all’attività

Avvio al campionamento probabilistico o casuale

Lo studio di un fenomeno, delimitato nel tempo e nello spazio, dovrebbe avvenire rilevando tutte le unità nelle quali quel fenomeno si manifesta, ossia l’intera **popolazione**. Ciò si verifica ad esempio quando si studia: la popolazione dei nati in un certo anno in un certo paese, quella dei matrimoni, l’andamento del prezzo di un certo bene ad intervalli temporali prefissati, e così via. Vi sono però situazioni in cui studiare la popolazione non è possibile ed è allora necessario limitarsi alla rilevazione di una parte della popolazione. Questa parte si chiama **campione**.

I motivi che giustificano la rilevazione e lo studio di un campione piuttosto che la rilevazione e lo studio dell’intera popolazione possono essere diversi:

- la popolazione può essere molto vasta. In tal caso può risultare troppo lungo analizzare tutte le unità statistiche (un sondaggio di opinioni viene fatto di solito su un insieme ristretto di individui poiché se ne vogliono conoscere gli esiti con rapidità);
- le misure possono essere distruttive; può essere il caso di un controllo delle misure di affidabilità di un componente elettronico;
- il costo della rilevazione totale della popolazione può essere esorbitante per le finalità della ricerca.

In tutti questi casi è necessario effettuare una indagine campionaria, provvedendo ad estrarre un campione dalla popolazione incognita che si intende indagare. Se l’estrazione è casuale si parla di **campione probabilistico** o **campione casuale,** in esso ogni elemento della popolazione ha una probabilità nota e non nulla di entrare a far parte del campione. Quando la probabilità di estrazione è la stessa per ogni elemento, il campione si dice **casuale semplice**. È chiaro così che campione casuale non significa campione preso “a casaccio”, ma campione che rispetta le leggi probabilistiche della tecnica di campionamento prescelto.

Di fronte ad un campione fornito da altri senza opportune indicazioni si potrebbe commettere l'errore di ritenerlo casuale quando invece non lo è, con ciò falsando l'interpretazione dei risultati dello studio. Questo errore è difficile da scoprire quando non si conoscono le modalità con le quali il campione è stato estratto. Ad esempio, se l'intento è di effettuare un sondaggio di opinioni sulle abitudini sportive degli abitanti di una città con età compresa tra 10 e 70 anni e si decide di scegliere un campione di 200 abitanti, esso non è casuale se le persone da intervistare vengono "prese":

- all'uscita da uno stadio;
- da un elenco telefonico (verrebbero esclusi tutti coloro che non compaiono nell'elenco);
- tra coloro che transitano su una strada (verrebbero esclusi i frettolosi, gli automobilisti, ecc. mentre si tenderebbe a raccogliere dati da coloro che sembrano più disposti a rispondere).

Per effettuare un campionamento realmente casuale si dovrebbe procedere nel seguente modo:

- si assegna ad ogni abitante della città tra 10 e 70 anni un numero progressivo;
- si pongono in un'urna tutti i numeri assegnati;
- si estraggono tanti numeri (in questo caso 200) quante sono le persone da intervistare;
- si intervistano le persone che corrispondono ai numeri estratti.

Procedendo in questo modo il campione estratto è effettivamente casuale, in quanto tutti i numeri dell'urna hanno la stessa probabilità di essere estratti (sono equiprobabili), e quindi tutti gli abitanti di quella città tra 10 e 70 anni hanno la stessa probabilità di essere intervistati.

Qualunque sia il modo in cui il campione viene estratto le informazioni da esso fornite vengono di fatto utilizzate per trarre delle conclusioni più generali, riferite a tutta la popolazione di cui il campione osservato è solo una parte, generalmente di numerosità molto minore. Ma la validità delle conclusioni tratte dipende dalla possibilità di controllare **l'errore di campionamento** (inteso come differenza fra l'informazione fornita dal campione con il "valore vero" nella popolazione), ciò è possibile solo se il campione è probabilistico. Il primo problema da affrontare nel trattare della teoria dei campioni è perciò quello di stabilire come il campione deve essere scelto in modo che si possa utilizzare la teoria della probabilità nell'effettuare il processo di inferenza che porta dal campione estratto alla popolazione incognita dalla quale l'estrazione è avvenuta.

Per tutelare le persone inesperte di fronte ai risultati dei sondaggi, l'Autorità per le Garanzie nelle Comunicazioni ha pubblicato il Testo del regolamento in materia di pubblicazione e diffusione dei sondaggi sui mezzi di comunicazione di massa (Delibera n. 237/03/CSP) che riporta gli elementi teorici di base che devono essere indicati perché i risultati presentati possano essere considerati significativi rispetto alla popolazione che si intende indagare.

Attività 1 - Il campionamento casuale e la stima della media di una popolazione

Fase 1 - Il problema

L'insegnante propone alla classe la seguente domanda: "Supponiamo di non poter misurare la statura di tutti gli ***N*** studenti della classe, se scegliamo a caso solo ***n*** studenti tra gli ***N*** e rileviamo le loro stature, è possibile che il campione ci fornisca informazioni attendibili sulla statura media degli studenti dell'intera classe?"

L'insegnante stimola una discussione al termine della quale emerge, per i motivi visti sopra, la necessità di un campionamento casuale; che possono portare alla formulazione di alcune domande del tipo:

a) quanto può essere grande ***n*** (***n*** numero degli elementi del campione, si indica con la lettera minuscola, per distinguerlo dal numero degli studenti dell'intera classe che indichiamo con ***N*** maiuscolo)?

b) come possono essere estratti gli ***n*** studenti?

c) nell'estrazione di un campione può capitare di estrarre più volte lo stesso studente?

d) due campioni sono da considerare diversi se lo stesso studente è estratto in posizioni diverse?

e) la media della statura calcolata sui dati di un campione (**media campionaria**) cambia al variare del campione?

f) la media del campione estratto (**media campionaria**) può essere usata come stima della statura media incognita della classe (**parametro della popolazione**)?

Il problema

Fase 2 - La popolazione e lo spazio dei campioni ordinati con reinserimento

Per aiutare gli studenti a rispondere a queste domande (e ad altre eventuali domande dello stesso tipo) l'insegnante propone una versione didatticamente più semplice ed efficace del problema prendendo in considerazione una popolazione di cinque studenti dello stesso genere (***N***=5) di cui si misura la statura.

Della popolazione si costruiscono gli indici di sintesi.

Studenti per statura (Popolazione)	
Studente	**Statura (cm)**
A	173
B	184
C	171
D	175
E	180

Tabella 1

Indici di sintesi della popolazione (parametri)		
Media	μ	176,6 (cm)
Varianza	σ^2	22,64 $(cm)^2$
Devst	σ	4,758 (cm)
Max	M	184 (cm)
Min	m	171 (cm)

Tabella 2

Utilizzando la Popolazione contenuta nella Tabella 1 l'insegnante fa costruire tutti i possibili campioni di due elementi (**n**=2), pervenendo alla Tabella 3. L'insieme di tutti i possibili campioni estratti da una popolazione si dice **spazio dei campioni** e dipende dalla tecnica di campionamento scelto.

Spazio dei campioni di 2 elementi scelti tra 5 (campionamento ordinato con reinserimento)					
Studente	A	B	C	D	E
A	(A,A)	(A,B)	(A,C)	(A,D)	(A,E)
B	(B,A)	(B,B)	(B,C)	(B,D)	(B,E)
C	(C,A)	(C,B)	(C,C)	(C,D)	(C,E)
D	(D,A)	(D,B)	(D,C)	(D,D)	(D,E)
E	(E,A)	(E,B)	(E,C)	(E,D)	(E,E)

Tabella 3

L'insegnante chiede cosa viene evidenziato dalla tabella 3.

Gli studenti, guidati dal docente, possono osservare che:

- i campioni ottenuti sono coppie ordinate scelte a partire dai cinque elementi della popolazione e di essi, i campioni scritti, rappresentano il prodotto cartesiano;
- sulla diagonale principale il campione è formato dalla stesso elemento ripetuto due volte;
- la matrice dei campioni non è simmetrica cioè il campione (A, B) è diverso dal campione (B, A);
- il numero di campioni ottenuto è 25 (25=5x5=52 , ossia **Nn**).

Nota per insegnante:
In Tabella 3 è agevole far verificare che ogni coppia estratta ha ugual probabilità di verificarsi ($25=5x5=5^2$, ossia N^n). L'insegnante potrà anche dire che qui si è effettuato un ***campionamento ordinato con ripetizione (anche detto bernoulliano)****.*

Poiché ad ogni coppia di studenti corrisponde una coppia di stature, l'insegnante invita gli studenti a costruire una nuova tabella che contiene in ogni casella le stature degli studenti che formano il corrispondente campione, ottenendo così:

Spazio dei campioni di 2 elementi scelti tra 5 (campionamento ordinato con reinserimento)					
Studente	A	B	C	D	E
A	(173;173)	(173;184)	(173;171)	(173;175)	(173;180)
B	(184;173)	(184;184)	(184;171)	(184;175)	(184;180)
C	(171;173)	(171;184)	(171;171)	(171;175)	(171;180)
D	(175;173)	(175;184)	(175;171)	(175;175)	(175;180)
E	(180;173)	(180;184)	(180;171)	(180;175)	(180;180)

Tabella 4

Come passo successivo, l'insegnante fa calcolare la media aritmetica delle stature di ciascuna coppia e la chiama **media campionaria** in quanto media aritmetica calcolata su un campione. La media contenuta in ogni casella è funzione del campione su cui è calcolata e si indica col simbolo $\bar{x}_i$ (qui i= 1,…, 25; in generale i= 1, …, N^n)

La media campionaria di ciascuno dei 25 campioni estratti viene riportata nella tabella seguente:

Media campionaria dei 25 campioni estratti					
	A	B	C	D	E
A	173	178,5	172	174	176,5
B	178,5	184	177,5	179,5	182
C	172	177,5	171	173	175,5
D	174	179,5	173	173	177,5
E	176,5	182	175,5	177,5	180

Tabella 5

Gli studenti osservano che le medie campionarie non sono tutte uguali da campione a campione, ma che il campione (A, B) ha la stessa media del campione (B, A), sicché la tabella 5 è simmetrica.

L'insieme di tutte le medie calcolate su tutti i campioni che compongono lo spazio dei campioni è una **variabile casuale (v.c.) campionaria**, o **stimatore.** La variabile casuale viene chiamata media campionaria ed indicata con il simbolo $\bar{X}$.

La Tabella 5 contiene perciò la variabile casuale media campionaria ottenuta dalla popolazione di Tabella 1, estraendo da essa tutti i campioni ordinati con ripetizione e di numerosità n=2.

Prima di passare alla fase successiva, come ulteriore rinforzo e verifica di quanto fin qui trovato, l'insegnante chiede alla classe:

- Qual è la differenza fra il concetto di parametro e quello di stima campionaria?
- Qual è la differenza fra media aritmetica della popolazione e media aritmetica campionaria?
- Quante medie aritmetiche della popolazione esistono? E quante medie aritmetiche campionarie?
- Due diversi campioni possono fornire medie aritmetiche campionarie diverse? E medie aritmetiche campionarie uguali?

Fase 3 - Analisi della distribuzione della v.c. media campionaria e suo utilizzo

L'insegnante ripropone la domanda con la quale si è aperta la fase 1 dell'attività e chiede agli studenti cosa è ora possibile rispondere:

"Supponiamo di non poter misurare la statura di tutti gli ***N*** studenti della classe, se scegliamo a caso solo ***n*** studenti tra gli ***N*** e rileviamo le loro stature, è possibile che il campione ci fornisca informazioni attendibili sulla statura media degli studenti dell'intera classe?"

Gli studenti dovrebbero essere in grado di comprendere che i campioni estraibili a caso, tenendo conto dell'ordine di estrazione e con reinserimento sono N^n. Per ogni campione è possibile calcolare la media aritmetica delle unità che lo compongono. Si tratta ora di chiedersi se e perché una qualsiasi di queste medie campionarie, non tutte diverse, possa essere utilizzata come stima della statura media della popolazione, ossia del parametro incognito che si vuole stimare e che si indica con μ.

L'insegnante chiede: un valore qualsiasi della tabella 5 - ad esempio 177,5 cm - può essere utilizzato come stima della media della classe?

Per aiutare nella risposta, l'insegnante invita gli studenti a costruire la distribuzione di frequenze delle medie campionarie ottenute nello spazio dei 25 campioni. I valori ottenuti sono presentati nella tabella 6.

Distribuzione della variabile casuale media campionaria nello spazio bernoulliano dei campioni di 2 elementi da una popolazione di 5		
$\bar{x}_i$ cm	n_i	f_i
171	1	1/25
172	2	2/25
173	3	3/25
174	2	2/25
175	1	1/25
175,5	2	2/25
176,5	2	2/25
177,5	4	4/25
178,5	2	2/25
179,5	2	2/25
180	1	1/25
182	2	2/25
184	1	1/25
Totale	25	1

Tabella 6

Nota per insegnante:
In tabella 6 è importante far osservare che la media campionaria descrive una variabile casuale i cui valori, compresi fra 171 e 184, sono riportati nella prima colonna. Nella terza colonna, per ogni valore è riportata la misura frequentista della sua probabilità, che coincide in questo caso con la definizione classica.

Visualizza il percorso didattico Stocastica e … legami intradisciplinari (http://forum.indire.it/repository_cms/working/export/5244/02_2.htm)

A questo punto l'insegnante fa calcolare la media, la varianza e la deviazione standard della distribuzione della variabile casuale media campionaria, il suo massimo e il suo minimo (Tabella 7) e li mette a confronto con i corrispondenti parametri della popolazione (Tabella 8).

$\bar{x}_i$ cm	n_i	f_i	$\bar{x}_i \cdot f_i$	$\bar{x}_i^2 \cdot f_i$
171	1	1/25	6,84	1169,64
172	2	2/25	13,76	2366,72
173	3	3/25	20,76	3591,48
174	2	2/25	13,92	2422,08
175	1	1/25	7	1225
175,5	2	2/25	14,04	2464,02
176,5	2	2/25	14,12	2492,18
177,5	4	4/25	28,4	5041
178,5	2	2/25	14,28	2548,98
179,5	2	2/25	14,36	2577,62
180	1	1/25	7,2	1296
182	2	2/25	14,56	2649,92
184	1	1/25	7,36	1354,24
Totale	25	1	176,6	31198,88

Media	176,6 (cm)
Varianza	11,32 (cm)^2
Devst	3,36 (cm)
Min	171 (cm)
Max	184 (cm)

Tabella 7

Indici di sintesi				
		popolazione		v.c. media campionaria
Media		176,6 (cm)		176,6 (cm)
Varianza		22,64 $(cm)^2$		11,32 $(cm)^2$
Devst		4,758 (cm)		3,36 (cm)
Min		171 (cm)		171 (cm)
Max		184 (cm)		184 (cm)

Tabella 8

Quali informazioni è possibile ricavare dagli indici di sintesi trovati nelle due distribuzioni?

La classe potrebbe notare che:

1. la media della distribuzione della v.c. media campionaria è uguale alla media delle stature nella popolazione;
2. la varianza della distribuzione della v.c. media campionaria è minore della varianza della popolazione. L'insegnante farà osservare che nella distribuzione delle medie campionarie la varianza è uguale a quella della popolazione diviso 2 (numerosità del campione);
3. il minimo e il massimo sono rispettivamente uguali 171 cm e 184 cm in entrambe le distribuzioni.

A seguito di quanto emerso l'insegnante potrebbe chiedere:

perché le due medie sono uguali? Ciò si verifica sempre?

Perché la varianza della v.c. media campionaria è minore di quella della popolazione? Ciò si verifica sempre? È possibile affermare che, quando si esegue un campionamento ordinato con ripetizione (o campionamento bernoulliano), la varianza della distribuzione delle medie campionarie è uguale alla varianza della popolazione diviso per la numerosità campionaria?

L'insegnante ribadisce che, per campioni casuali semplici ordinati con reinserimento di numerosità n, valgono le seguenti proprietà:

- il valor medio della distribuzione della v.c. media campionaria è uguale alla media della popolazione;
- la varianza della v.c. media campionaria è pari al rapporto fra la varianza della popolazione e la numerosità del campione.

Nota per insegnante:

Le proprietà enunciate si possono esprimere in simboli nel modo seguente: $M(\overline{X})=\mu$ *e* $Var(\overline{X})=\frac{\sigma^2}{n}$

$$M[\overline{X}]=M\left[\frac{\sum_{i=1}^{n}X_i}{n}\right]=\frac{1}{n}\sum_{i=1}^{n}M[X_i]=\frac{1}{n}\cdot n\cdot\mu=\mu \quad Var(\overline{X})=Var\left[\frac{\sum_{i=1}^{n}X_i}{n}\right]=\frac{1}{n^2}\sum_{i=1}^{n}Var[X_i]=\frac{1}{n^2}\cdot n\cdot\sigma^2=\frac{\sigma^2}{n}$$

Dove Xi è la v.c. generata dall'i-esimo elemento estratto del campione. Nell'universo bernoulliano dei campioni le variabili Xi (i=1, …, N^n*) hanno tutte la stessa distribuzione di probabilità con media e variabilità uguali a quelle della variabile da cui viene estratto il campione; le v.c. Xi sono per costruzione fra loro indipendenti. Quindi hanno la stessa media e la stessa varianza e pari a quelle della popolazione di partenza. (Si confronti, ad esempio, con Tabella 4)*

I simboli introdotti rappresentano rispettivamente:

M	*Operatore valor medio*
Var	*Operatore varianza*
$\overline{X}$	*Variabile casuale Media campionaria*
μ	*Media della popolazione*
σ^2	*Varianza della popolazione*

Va anche osservato che, nota la distribuzione della v.c. media campionaria per l'universo dei campioni, il calcolo della sua media e della sua varianza è dato da:

$$\mu = \frac{\sum_{i=1}^{k} \bar{x}_i \cdot n_i}{\sum_{i=1}^{k} n_i} = \sum_{i=1}^{k} \bar{x}_i \cdot f_i$$

$$Var(\bar{X}) = \frac{\sum_{i=1}^{k} (\bar{x}_i - \mu)^2 \cdot n_i}{\sum_{i=1}^{k} n_i} = \sum_{i=1}^{k} (\bar{x}_i - \mu)^2 \cdot f_i = \sum_{i=1}^{k} (\bar{x}_i)^2 \cdot f_i - \mu^2$$

I calcoli relativi all'uso delle formule di cui sopra sono riportati in tabella 7.

L'aver trovato che la distribuzione della v.c. media campionaria ha media uguale a quella della popolazione (parametro da stimare) e varianza pari alla varianza della popolazione divisa per la numerosità del campione ci porta ad affermare che una qualsiasi media campionaria appartiene ad un insieme la cui media è uguale alla media della popolazione (μ) e che tale stima ha in media una distanza da μ (in termini di deviazione standard) pari a σ/√n. E quindi, se non si conosce la media della popolazione, è possibile stimarla con uno dei valori assunti dalla v.c. media campionaria nel campione ordinato con ripetizione e perciò anche con 177,5 cm, resta inteso che in media questa stima ha un errore campionario che in media è uguale a 3,36 cm.

Nota per insegnante:

Il cammino logico intrapreso è senza dubbio complesso e può essere utile che l'insegnante sottolinei esplicitamente ciò che gli studenti devono avere appreso.

*Occorre distinguere fra un parametro nella popolazione - che è un **numero -** e che generalmente è incognito, e lo stimatore campionario -che è una **v.c**.*

*La **popolazione** è formata da tutte le unità (**N**) nelle quali si manifesta il fenomeno che si intende studiare. Se si studia un carattere quantitativo la popolazione ha indici di sintesi che sono: la media aritmetica (μ), la varianza (σ) e la deviazione standard (σ), il minimo e il massimo. Si tratta comunque di numeri!!!.*

*Se non è possibile studiare la popolazione, per stimare i suoi parametri si può utilizzare un campione di numerosità **n**, generalmente molto minore di **N**. L'applicazione di uno stimatore campionario (funzione degli elementi del* *campione estratto) ad un campione estratto produce una* **stima campionaria***, sicché, se lo stimatore è la media aritmetica, sul campione si calcola la stima campionaria* $\overline{x}_i$.

Adottando lo schema di campionamento probabilistico con ripetizione nel quale ha importanza l'ordine di estrazione, si ottiene lo spazio dei campioni equiprobabili contenente $\boldsymbol{N^n}$ *elementi. Ammesso che interessi stimare μ, in questo spazio, sostituendo ad ogni campione la sua media si ottiene la distribuzione della variabile casuale media campionaria. Come per qualsiasi distribuzione anche per essa si può calcolare la media e la varianza. La v.c. media campionaria gode di due proprietà fondamentali per il ragionamento statistico inferenziale induttivo: utilizzando il campionamento casuale semplice con ripetizione in cui ha importanza l'ordine si ha che qualunque campione ottenuto appartiene ad un insieme la cui media riproduce il parametro che si intende stimare, e la cui varianza è tanto minore quanto maggiore è n, la numerosità campionaria. Sicché al crescere di n, il campione diventa grande e l'errore di campionamento, che si commette usando la stima campionaria per prevedere il parametro della popolazione (μ), in media si riduce.*

Attività 2 - Forma della distribuzione della variabile casuale media campionaria

Fase 1- Simulazione della distribuzione delle medie campionarie. Una nuova popolazione e un sottospazio dello spazio campionario

Lo studio di ogni distribuzione statistica include oltre al calcolo delle misure di sintesi, anche la costruzione e il commento della sua rappresentazione grafica, per coglierne l'andamento. L'insegnante propone allora alla classe di rappresentare con un grafico la v. c. di **Tabella 6**.

Distribuzione della variabile casuale media campionaria nello spazio bernoulliano dei campioni di 2 elementi da una popolazione di 5		
$\bar{x}_i$ cm	n_i	f_i
171	1	1/25
172	2	2/25
173	3	3/25
174	2	2/25
175	1	1/25
175,5	2	2/25
176,5	2	2/25
177,5	4	4/25
178,5	2	2/25
179,5	2	2/25
180	1	1/25
182	2	2/25
184	1	1/25
Totale	25	1

Tabella 6

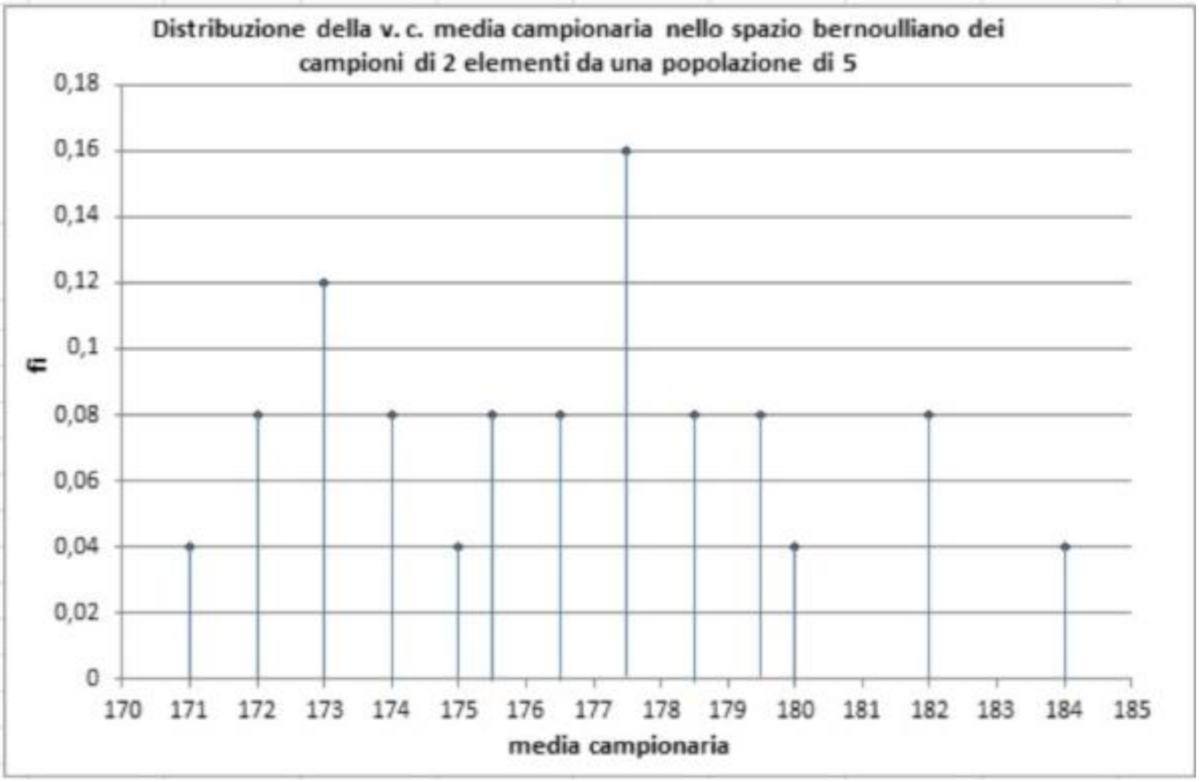


Figura 1

L'insegnante fa osservare che immaginare di avere a che fare con una popolazione di 5 unità se ha fatto emergere alcune considerazioni fondamentali sullo spazio dei campioni, sulla costruzione di una variabile casuale, sui suoi valori di sintesi confrontati coi parametri della popolazione, non è in grado di cogliere l'andamento generale della variabile casuale.

In effetti il grafico di Figura 1 fornisce diverse informazioni utili: si tratta di una distribuzione con una moda in 177,5 cm, per la quale il valore minimo e il massimo coincidono rispettivamente con quelli della popolazione, la distribuzione appare sgranata e con un andamento altalenante sulle code; tuttavia questo grafico dà elementi informativi limitati sul modello che esprime l'andamento della variabile casuale media campionaria.

Se si vuole cogliere l'andamento della variabile casuale l'insegnante propone di cambiare popolazione e di riferirsi ad un esempio più realistico. Suggerisce di studiare sperimentalmente, col metodo della simulazione, la distribuzione della media campionaria delle stature calcolate su campioni di numerosità *n=5* estratti dagli *N=18* studenti della classe (i numeri 5 e 18 non sono prescrittivi, ovviamente).

Si riportano, in Tabella 9, le stature dei 18 studenti (popolazione di riferimento), la media, la varianza e la deviazione standard, il minimo e il massimo delle stature.

Studenti della classe 5^ ISB per statura (popolazione)	
Studente	statura (cm)
1	173
2	184
3	171
4	175
5	180
6	176
7	193
8	170
9	186
10	187
11	164
12	183
13	178
14	177
15	180
16	194
17	197
18	188

Tabella 9

Indici di sintesi della popolazione		
Media	μ	180,89 (cm)
Varianza	σ^2	75,21 $(cm)^2$
Devst	σ	8,67 (cm)
Max	M	197 (cm)
Min	m	164 (cm)

Tabella 10

L'insegnante, proposto di costruire campioni di 5 elementi scelti casualmente tra i 18 studenti, chiede alla classe se sia possibile utilizzare il metodo adottato in precedenza per costruire lo spazio dei campioni.

- Da quanti campioni è formato lo spazio bernuolliano dei campioni?
- Si riesce fattivamente a costruirlo?
- Può valere la pena costruirne solo un sottospazio? Hanno mai sentito la parola simulazione?
- Cosa significa?
- Il computer può essere di aiuto?

Gli studenti di fronte all'evidenza che costruire i 18^5 campioni che formano lo spazio campionario è un'impresa che non vale lo sforzo fatto, prenderanno in considerazione la proposta di farne una simulazione al computer, usando un foglio elettronico e la relativa funzione che restituisce un numero intero casuale maggiore di 0 e minore od uguale a 18 e legando ad ogni estrazione fatta dal computer l'altezza dello studente corrispondente.
L'insegnante suggerisce di costruire col computer 5000 campioni.
Scarica il file simulazioni [*file formato. xls*]

Di seguito si riporta la distribuzione delle medie campionarie ottenute dalla simulazione di 5000 campioni di numerosità 5 estratti, facendo uso di un foglio di calcolo, dalla popolazione di tabella 9. Tale popolazione si ricorda ha media: μ=180,89 cm e varianza: σ^2 = 75,21 cm^2.

L'insegnante fa osservare alla classe che l'elevato numero di campioni generati con il computer (5000) porta all'esigenza di raggruppare in classi i valori delle medie campionarie ottenute. La scelta dell'ampiezza delle classi va fatta in modo che si possa ottenere un grafico in grado di esprimere con sufficiente chiarezza l'andamento della distribuzione. Il gruppo di studenti che ha costruito la tabella ha deciso di fare classi di ampiezza 2 visto che nelle varie simulazioni provate questa era la scelta più adatta fra le proposte emerse.

Nota per insegnante:
Il raggruppamento in classe è necessario quando il numero delle informazioni è elevato e i dati sono molto densi. Un numero troppo basso di classi, raggruppando eccessivamente i dati, determina una perdita di informazione sull'andamento della distribuzione e la rende non significativa; è intuitivo che una o due sole classi determinano l'impossibilità di evidenziare qualunque caratteristica della distribuzione. Inversamente, ma con un risultato finale simile, un numero troppo elevato di classi disperde i valori e non rende manifesta la forma della distribuzione.
È opportuno che il numero delle classi non sia troppo "grande", che l'ampiezza delle stesse non sia troppo "grande" e che la frequenza ad esse associata non presenti valori troppo piccoli. Per suddividere i dati osservati in classi è possibile utilizzare una delle seguenti proposte:
a) fissare l'ampiezza della classe e quindi individuare il numero di classi in cui suddividere il collettivo osservato: $\textit{numero classi} = \dfrac{Max - \min}{\textit{ampiezza classe}}$
b) fissare il numero di classi e individuare l'ampiezza costante delle classi: $\textit{ampiezza classe} = \dfrac{Max - \min}{\textit{numero classi}}$

Di seguito si riportano la distribuzione di frequenze della v.c. media campionaria e la tabella di confronto tra gli indici di sintesi della variabile campionaria simulata e i corrispondenti parametri nella popolazione.

Distribuzione di frequenza delle medie campionarie per campioni di numerosità 5 (distribuzione simulata di 5000 campioni)		
$\bar{x}_i$ cm	n_i	f_i
162--|164	0	0
164--|166	0	0
166--|168	1	0,0002
168--|170	6	0,0012
170--|172	35	0,007
172--|174	161	0,0322
174--|176	368	0,0736
176--|178	618	0,1236
178--|180	924	0,1848
180--|182	972	0,1944
182--|184	860	0,172
184--|186	581	0,1162
186--|188	300	0,06
188--|190	125	0,025
190--|192	45	0,009
192--|194	4	0,0008
194--|196	0	0
196--|198	0	0
Totale	5000	1

Tabella 11

Indici di sintesi			
		popolazione	media campionaria simulata rispetto a 5000 campioni
Media	μ	180,89 (cm)	180,82 (cm)
Varianza	σ^2	75,21 $(cm)^2$	15,63 $(cm)^2$
Devst	σ	8,67 (cm)	3,95 (cm)

Tabella 12

L'insegnante fa notare che le caratteristiche emerse nell'attività 1, riferite all'intero spazio dei campioni, emergono solo tendenzialmente anche in questa attività che prevede lo studio della v.c. media campionaria su un numero limitato di campioni prodotti da una simulazione.

L'insegnante stimola la discussione ponendo domande del tipo:

a) come mai in tabella 11 ci sono classi che non hanno frequenze?

b) come mai in tabella 12 il valore medio della popolazione e della distribuzione della media campionaria sono diversi mentre gli stessi indici per l'universo dei campioni di due elementi della attività 1 erano risultati uguali?

c) l'insegnante fa calcolare il rapporto tra la varianza nota della popolazione e la numerosità del campione e chiede se il valore trovato è uguale a quello della varianza della distribuzione della media campionaria riportato in tabella 12;

d) se la stima di un campione estratto è 181,6 (cm) quale errore si commette utilizzandola per prevedere la media della popolazione μ= 180,89 (cm)? Volendo stabilire se l'errore è piccolo o grande quale termine di confronto è appropriato?

Dalla discussione dovrà emergere emerge che: per il punto a) i 5000 campioni generati sono solo alcuni dell'universo dei possibili campioni casuali di 5 elementi scelti, con reinserimento, tra 18 e quindi è possibile che nella simulazione manchino alcuni valori medi.

Per il punto b) si evidenzia che i 5000 campioni scelti sono solo una parte dell'universo dei campioni mentre **nell'attività 1** la distribuzione delle medie campionarie fa riferimento all'intero universo dei **campioni**.

Per il punto c) anche la motivazione sulla diversità della varianza è spiegata dal fatto che la distribuzione utilizzata è solo la simulazione di una parte dello spazio dei campioni, ciò comporta fra l'altro che la varianza della media campionaria simulata è stata calcolata con riferimento agli scarti da 180,82 cm e non da μ = 180,89, come si sarebbe verificato nello spazio dei 18^5 campioni possibili.

Ma se la numerosità dello spazio dei campioni è così importante, quanti campioni formano nel nostro caso lo spazio dei campioni? La risposta è 18^5 = 1.889.568 ovvero circa 400 volte la numerosità del campione scelto.

Nota per insegnante:

È opportuno che l'insegnante si assicuri che gli studenti sappiano calcolare 18^5 *potranno usare i logaritmi, la calcolatrice tascabile, il computer.*

Si suggerisce di sollecitare gli studenti a fare commenti sui risultati ottenuti.

Per il punto d) l'insegnante fa notare che 181,6 (cm) è uno dei valori possibili della variabile media campionaria la cui media è LA MEDIA DELLA POPOLAZIONE.

Ricordando il significato della deviazione standard in una distribuzione (che misura quanto mediamente i dati si discostano dalla loro media) è possibile affermare che l'errore che si commette prendendo 181,6 cm (oppure un qualsiasi altro valore della media campionaria simulata) come valore "vero" della media incognita della popolazione è stimabile attraverso il valore della deviazione standard della distribuzione simulata delle medie campionarie e quindi esso dista mediamente dal valor medio della distribuzione simulata 3,95 cm, che è una stima di $\sigma/\sqrt{n}$ = 3,88cm.

Nota per insegnante:

Nel caso in cui non sia possibile disporre di un calcolatore, si può sviluppare l'attività proponendo ad ogni studente di estrarre 5 campioni, con reinserimento di scrivere la statura misurata su ciascuna unità e di calcolare la media di ogni campione ottenuto. Le medie campionarie trovate dagli studenti saranno nel loro insieme utilizzate per: simulare la distribuzione di frequenze delle medie campionarie ottenute da una popolazione di media nota: μ=180,89 cm e varianza pure nota: σ^2 = 75,21 cm^2; calcolare i suoi indici di sintesi e fare le considerazioni opportune.

Resta inteso che la simulazione darà risultati tanto migliori quanto più il numero dei campioni estratti sarà elevato.

Fase 2 - Rappresentazione grafica della distribuzione simulata

L'insegnante invita la classe a rappresentare, tramite istogramma, la distribuzione delle medie campionarie e a commentarne l'andamento. L'insegnante ricorda che per costruire l'istogramma si deve calcolare la densità di frequenza ($h_i = \frac{f_i}{x_{i+1} - x_i}$).

Visualizza il percorso didattico I giovani e la musica (http://forum.indire.it/repository/working/export/5101/fase7.htm): rappresentazioni grafiche.

Di seguito si riporta il grafico della distribuzione presente in tabella 11 con le densità di frequenza sull'asse delle ordinate.

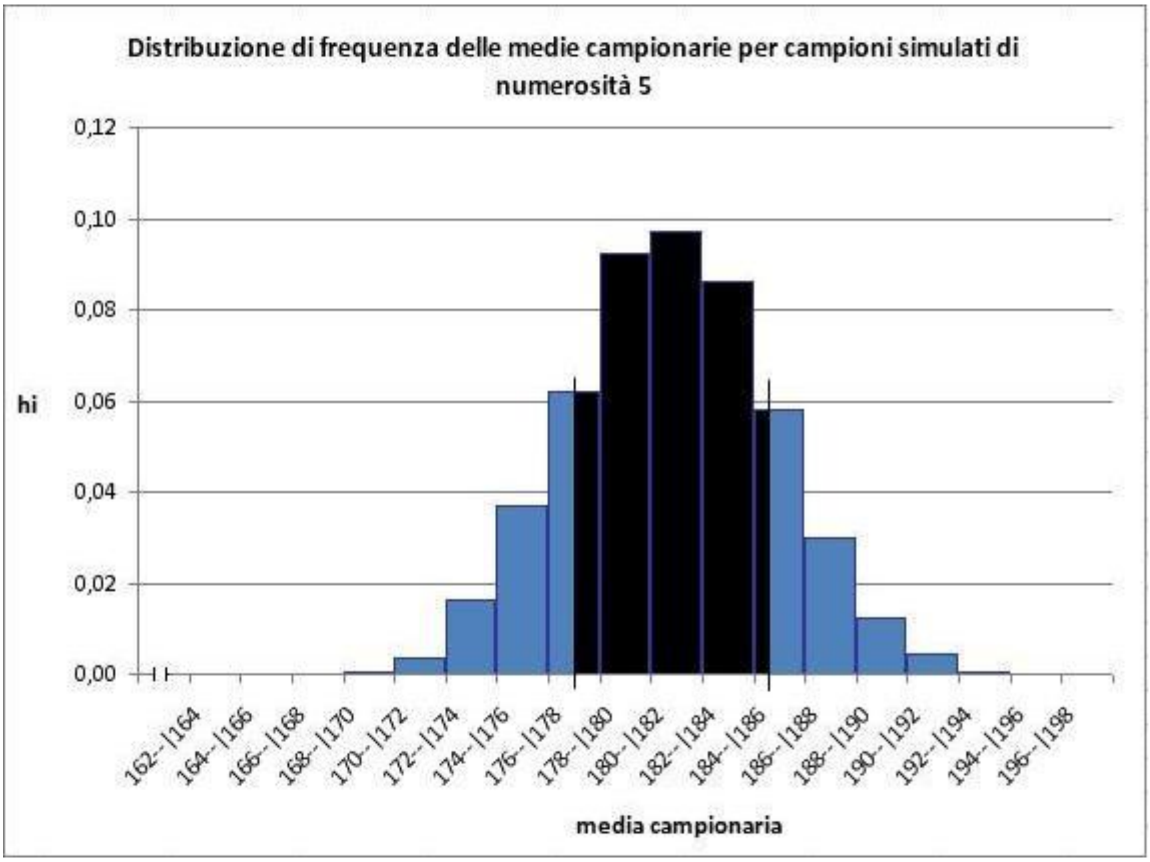


Figura 2

Nota per insegnante:

Può essere opportuno ricordare alla classe che:

- *l'area racchiusa sotto la spezzata rappresenta la somma totale delle frequenze relative e quindi vale 1;*
- *ciascun rettangolo ha un'area proporzionale alla frequenza relativa associata alla classe la cui ampiezza corrisponde alla base;*
- *le frequenze relative in questo contesto sono misure di probabilità, e dunque in figura 2 le probabilità sono misure di aree.*

La discussione sulla Figura 2 porta ad osservare che: la distribuzione è tendenzialmente simmetrica rispetto alla classe che contiene il valor medio

(180−|182), inoltre questa è anche la classe con maggiore densità ed è perciò la classe modale. L'insegnante chiede:

nella distribuzione ottenuta per simulazione qual è la probabilità che venga estratto a caso un campione che fornisce una stima della media ($\bar{X}$ i):

- compresa nell'intervallo $\mu \pm \frac{\sigma}{\sqrt{n}}$, ovvero, qual è la probabilità che una media campionaria sia compresa nell'intervallo [177,01; 184,77]?
- compresa nell'intervallo $\mu \pm \frac{2 \cdot \sigma}{\sqrt{n}}$ [173,14; 188,64]?
- sia compresa nell'intervallo $\mu \pm \frac{3 \cdot \sigma}{\sqrt{n}}$ [169,26; 192,52]?L'insegnante ricorda agli studenti che si deve calcolare l'area della parte dell'istogramma compresa nell'intervallo definito ponendo attenzione al caso in cui gli estremi dell'intervallo siano interni ad una classe. Per il primo caso si riportano i calcoli effettuati aiutandosi con i dati della tabella 13 dove sono state riportate anche le densità di frequenza ed, eventualmente, con la rappresentazione grafica di figura 3 dove è evidenziata in colore diverso la zona interessata racchiusa tra i valori [177,01 ; 184,77].

Distribuzione di frequenza delle medie campionarie per campioni di numerosità 5 (distribuzione simulata di 5000 campioni)			
$\bar{x}_i$ cm	n_i	f_i	hi
162--|164	0	0	0
164--|166	0	0	0
166--|168	1	0,0002	0,0001
168--|170	6	0,0012	0,0006
170--|172	35	0,007	0,0035
172--|174	161	0,0322	0,0161
174--|176	368	0,0736	0,0368
176--|178	618	0,1236	0,0618
178--|180	924	0,1848	0,0924
180--|182	972	0,1944	0,0972
182--|184	860	0,172	0,0860
184--|186	581	0,1162	0,0581
186--|188	300	0,06	0,0300
188--|190	125	0,025	0,0125
190--|192	45	0,009	0,0045
192--|194	4	0,0008	0,0004
194--|196	0	0	0
196--|198	0	0	0
Totale	5000	1	

Tabella 13

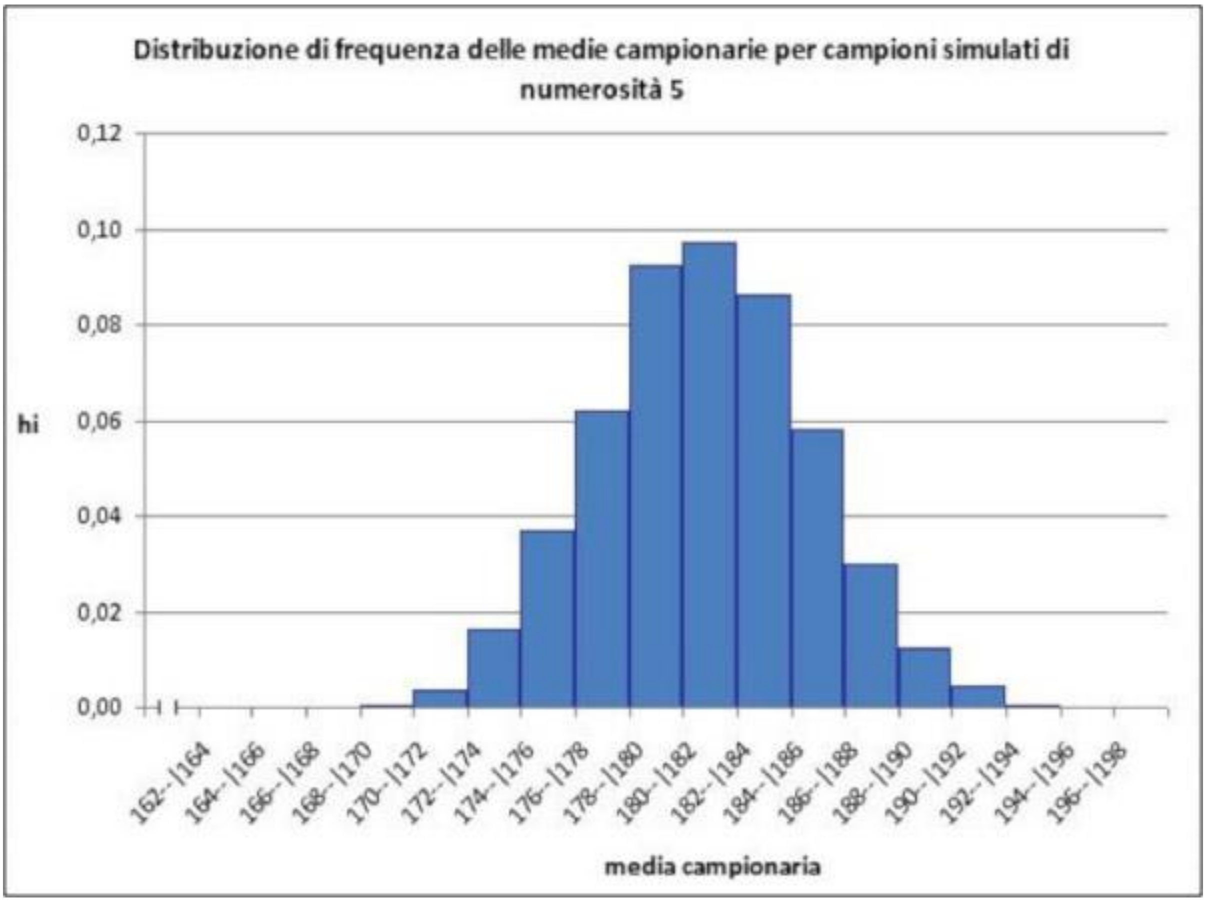


Figura 3

Area annerita = Prob [177,01≤ $\bar{x}_i$ ≤184,77]= (178 - 177,01)*0,0618+0,1848+0,1944+0,172+(184,77-184)*0,0581=0,6571

Gli studenti trovano, usando un procedimento analogo, che per il secondo intervallo la probabilità richiesta è pari:

(174-173,14)*0,0161+0,0736+0,1236+0,1848+0,1944+0,172+0,1162+0,06+(188,64-188)*0,0125= 0,9576

e che per il terzo intervallo la probabilità richiesta è pari a 0,9984.

Dopo avere messo in evidenza la forma della distribuzione simulata e aver calcolato su di essa alcune probabilità importanti, facendo venire meno l'ipotesi fondamentale su cui si sono sviluppate fino ad ora considerazioni di inferenza deduttiva, e cioè che la popolazione di partenza sia completamente nota, l'insegnante avvia gli studenti ai primi rudimenti del ragionamento inferenziale induttivo.

L'insegnante propone esplicitamente la nuova ipotesi di lavoro: "Non si conosce la popolazione, ma si ha a disposizione la simulazione della distribuzione della variabile casuale media campionaria" e mostra il grafico di Figura 1. In questa nuova situazione, chiede l'insegnante:

- se avendo estratto un campione casuale con ripetizione da una popolazione di media incognita μ, ne avessimo calcolato la media $\bar{x}_i$ potremmo dirci certi che $\bar{x}_i$=μ?
- Potremmo dirci certi che $\bar{x}_i$ è la manifestazione di una v.c. la cui media è uguale a μ?
- Prima di estrarre un campione casuale con ripetizione di dimensione 5 dalla popolazione di tabella 9, secondo voi, quale zona del grafico di Figura 1 avrà la maggiore probabilità di contenere la media μ con un errore quadratico medio pari a $\pm\frac{2\cdot\sigma}{\sqrt{n}}$?
- Secondo voi è maggiore la probabilità di commettere un errore di ampiezza $\pm\frac{2\cdot\sigma}{\sqrt{n}}$ o un errore di ampiezza $\pm\frac{\sigma}{\sqrt{n}}$?

Dalla discussione le risposte date dagli studenti potrebbero essere:

- no, in quanto al variare del campione il valore della media campionaria può variare come osservato nella prima attività svolta e perciò non è detto che nel campione estratto si ottenga la relazione $\bar{x}_i$=μ; dunque si tratta di un evento probabile, ma non certo;
- sì come osservato dalle due attività proposte e dalle osservazioni fatte per esse;
- da quanto detto in precedenza si deduce che si tratta dell'intervallo, simmetrico rispetto alla media, di ampiezza pari alla differenza tra $\mu+\frac{2\cdot\sigma}{\sqrt{n}}$ e $\mu-\frac{2\cdot\sigma}{\sqrt{n}}$;

- dai calcoli precedenti e dalla Figura 2 la risposta é ovvia, essendo il primo valore 0,9576 e l'altro 0,6571.

Nota per insegnante:
La riposta alla terza domanda può, a questo livello, essere ancora imprecisa, l'importante è che gli studenti si orientino ad individuare un intervallo sull'asse delle ascisse, intorno al valore centrale con una distanza circa doppia rispetto a quella esistente fra il valore medio e l'ascissa del punto di flesso.

Indicazioni metodologiche

Questa attività può essere introdotta in una classe quarta della scuola secondaria di secondo grado dopo che l'insegnante ha verificato le conoscenze di base sulle nozioni della statistica descrittiva e di calcolo delle probabilità.
È opportuno che l'insegnante aiuti gli studenti a capire la necessità di effettuare un campionamento o per ottenere informazioni in modo rapido o perché l'informazione richiesta non può essere altro che una informazione campionaria. Nella costruzione di semplici spazi (o universi) dei campioni casuali gli studenti possono lavorare in piccoli gruppi, mettendo a confronto i diversi approcci e le diverse tecniche campionarie utilizzate. Seguirà una discussione e un confronto collettivo per arrivare ad una formalizzazione, da parte dell'insegnante, dei concetti emersi dall'attività compresa la necessità di conoscere le nozioni base del calcolo combinatorio per la costruzione dell'universo dei campioni e per la soluzione di problemi che richiedono la costruzione di associazioni fra elementi ed il loro conteggio.
È in ogni caso opportuno che le situazioni più complicate, che richiedono di trovare uno spazio campionario di numerosità elevata, vengano risolte attraverso l'uso del computer. Questa modalità didattica coinvolge emotivamente gli studenti, li rassicura e può facilitare l'apprendimento di questo delicato e fondamentale argomento.

Spunti per approfondire

Approfondimenti disciplinari

Attività 1 - Campionamento ordinato senza reinserimento

L'insegnante propone alla classe di modificare lo schema di campionamento dell'attività 1, in particolare propone di evitare la ripetizione dello stesso elemento. In concreto questo significa togliere dalla tabella 4 (che si trova nell'Attività 1, fase 2) i campioni che stanno sulla diagonale principale e dunque tutte le medie campionarie corrispondenti: ciò implica che, ad esempio, in un sondaggio, si intervisti due volte la stessa persone. Chiede perciò agli studenti di identificare il nuovo spazio campionario e fa loro costruire la nuova distribuzione della media campionaria.

Media campionaria dei campioni estratti escludendo quelli riferiti allo stesso studente					
	A	B	C	D	E
A		178,5	172	174	176,5
B	178,5		177,5	179,5	182
C	172	177,5		173	175,5
D	174	179,5	173		177,5
E	176,5	182	175,5	177,5	

Tabella 14

Distribuzione della media campionaria delle stature per campioni di 2 elementi da una popolazione di 5, senza ripetizione		
$\bar{x}_i$ cm	n_i	f_i
172	2	1/10
173	2	1/10
174	2	1/10
175,5	2	1/10
176,5	2	1/10
177,5	4	1/5
178,5	2	1/10
179,5	2	1/10
182	2	1/10
Totale	20	1

Tabella 15

A questo punto l'insegnante fa calcolare la media, la varianza e la deviazione standard della nuova distribuzione delle medie campionarie, che ora varia fra 172 e 182, e li mette a confronto con quelli della popolazione.

$\bar{x}_i$ cm	n_i	f_i	$\bar{x}_i \cdot f_i$	$\bar{x}_i^2 \cdot f_i$
172	2	1/10	17,2	2958,40
173	2	1/10	17,3	2992,90
174	2	1/10	17,4	3027,60
175,5	2	1/10	17,55	3080,03
176,5	2	1/10	17,65	3115,23
177,5	4	1/5	35,5	6301,25
178,5	2	1/10	17,85	3186,23
179,5	2	1/10	17,95	3222,03
182	2	1/10	18,2	3312,40
Totale	20	1	176,6	31196,05
			Media	176,6 (cm)
			Varianza	8,49 (cm)^2
			Devst	2,91 (cm)
			Min	172 (cm)
			Max	182 (cm)

Tabella 16

Indici di sintesi				
		popolazione		media campionaria
Media		176,6 (cm)		176,6 (cm)
Varianza		22,64 $(cm)^2$		8,49 $(cm)^2$
Devst		4,758 (cm)		2,91 (cm)
Min		171 (cm)		172 (cm)
Max		184 (cm)		182 (cm)

Tabella 17

Gli studenti osservano che, anche se il numero dei campioni è diminuito, la media della distribuzione delle medie campionarie è rimasta uguale alla media della popolazione mentre la varianza è diminuita rispetto al precedente spazio dei campioni.

L’insegnante fa riflettere che è cambiata la tecnica di campionamento e che le medie campionarie generate senza ripetizione sono “più vicine” alla media della popolazione.

Attività 2 - Campionamento senza reinserimento e senza ordine o campionamento in blocco

Come ulteriore schema di campionamento l'insegnante propone di eliminare dalla **tabella 4** anche i campioni che si trovano nella parte superiore della stessa, e di fare costruire la distribuzione della media campionaria collegata al nuovo spazio dei campioni. Fa notare che questa scelta implica di "non tener conto dell'ordine in cui si sono scelti gli studenti".

Media campionaria dei campioni estratti escludendo sia quelli riferiti allo stesso studente sia quelli con elementi uguali ma diversi per l'ordine					
	A	B	C	D	E
B	178,5				
C	172	177,5			
D	174	179,5	173		
E	176,5	182	175,5	177,5	

Tabella 18

La distribuzione delle media campionarie in questo caso è riportata nelle tabelle che seguono assieme ai valori di sintesi:

Distribuzione della media campionaria delle stature per campioni di 2 elementi da una popolazione di 5, senza ripetizione e senza ordine		
$\bar{x}_i$ cm	n_i	f_i
172	1	1/10
173	1	1/10
174	1	1/10
175,5	1	1/10
176,5	1	1/10
177,5	2	1/5
178,5	1	1/10
179,5	1	1/10
182	1	1/10
Totale	10	1

Tabella 19

$\bar{x}_i$ cm	n_i	f_i	$\bar{x}_i \cdot f_i$	$\bar{x}_i^2 \cdot f_i$
172	1	1/10	17,2	2958,40
173	1	1/10	17,3	2992,90
174	1	1/10	17,4	3027,60
175,5	1	1/10	17,55	3080,03
176,5	1	1/10	17,65	3115,23
177,5	2	1/5	35,5	6301,25
178,5	1	1/10	17,85	3186,23
179,5	1	1/10	17,95	3222,03
182	1	1/10	18,2	3312,40
Totale	10	1	176,6	31196,05

Media	176,6 (cm)
Varianza	8,49 (cm)^2
Devst	2,91 (cm)
Min	172 (cm)
max	182 (cm)

Tabella 20

L’insegnante fa notare che nella Attività 1 e Attività 2 si sono ottenuti, in termini di indici di sintesi della distribuzione delle medie campionarie, gli stessi risultati e quindi la scelta di “non reinserire l’elemento scelto” e “tener conto dell’ordine”

oppure di “non reinserire l’elemento scelto” e “non tener conto dell’ordine” non modifica l’informazione richiesta sul valor medio della popolazione.
Quale delle tecniche campionarie conviene utilizzare per ottenere stime “migliori” della media della popolazione?
La discussione in classe porta gli studenti ad osservare che se le medie delle distribuzioni sono rimaste uguali nei diversi tipi di campionamento, quando il campionamento è senza ripetizione la varianza è diminuita. Questo porta a dire che i valori della variabile casuale media campionaria sono, mediamente, più vicini alla media della popolazione quando i campioni sono senza ripetizione. e quindi, per popolazioni di numerosità piccola, sarebbe più conveniente utilizzare la tecnica di campionamento senza reinserimento.

Spunti per altre attività con gli studenti

Il calcolo combinatorio

L'insegnante propone alla classe alcune riflessioni sul procedimento utilizzato per la creazione degli spazi dei campioni nelle attività che riguardano il campionamento osservando che si sono utilizzati campioni di numerosità piccola (2 scelti tra cinque; 5 scelti tra 18) con tecniche diverse (ordinati e con ripetizione, ordinati senza ripetizione, non ordinati senza ripetizione). Da queste riflessioni l'insegnante presenta, partendo da una serie di problemi, anche le principali nozioni sul calcolo combinatorio visto come strumento per risolvere problematiche legate sia alle tecniche campionarie che alla scelta ed enumerazione di associazioni di elementi in contesti diversi.

Nota per insegnante:
Come tecnica di campionamento non è stata presentata quella che prevede estrazioni di campioni non ordinati e con ripetizione in quanto non utilizzata nelle indagini campionarie

Fase 1

Principio fondamentale del calcolo combinatorio

L'insegnante, dopo aver diviso la classe in gruppi propone a ciascun gruppo una delle seguenti situazioni:

Situazione 1: Una bicicletta sportiva monta una meccanica con tre corone dentate anteriori e un pignone posteriore con un numero di ingranaggi pari a sette. In quanti modi un ciclista può scegliere a caso il rapporto con cui correre?

Situazione 2: Ogni mattina Alessandra sceglie, a caso, l'abbigliamento per andare a scuola. Se nel suo armadio ci sono quattro paia di pantaloni e cinque felpe, in quanti modi diversi può vestirsi indossando a caso un pantalone e una felpa?

Situazione 3: Il menù di un ristorante mette a disposizione della clientela 5 antipasti, 3 primi, 4 secondi piatti e 3 varietà di dolci. In quante maniere è possibile organizzare un pasto, scegliendo a caso un antipasto, un primo, un secondo ed un dolce?

Gli studenti dovrebbero osservare che problemi di questo tipo possono essere risolti utilizzando il **prodotto cartesiano** dei dati del problema.

Pertanto la risposta ai tre problemi presentati è:

per il problema 1

3 x 7 = 21 rapporti

per il problema 2

4 x 5 = 20 abbinamenti di pantaloni e felpa

per il problema 3

5 x 3 x 4 x 3 = 180 pasti diversi.

A questo punto l'insegnante potrebbe formalizzare la soluzione fornendo la seguente regola:

se un evento E_1 può accadere in *p* modi diversi, un evento E_2 può accadere in *q* modi, un evento E_3 può accadere in *r* modi, …., allora il *numero* totale dei modi in cui accade l'evento E_1, seguito dalle evento E_2 e poi dall'evento E_3 ecc, uno indipendentemente dall'altro, è dato da:

N = $p \times q \times r$ ….

Chiama questa regola come **il principio fondamentale del calcolo combinatorio**.

Fase 2

Disposizioni semplici

L'insegnante propone, alla classe divisa in gruppi, una successiva serie di situazioni problematiche:

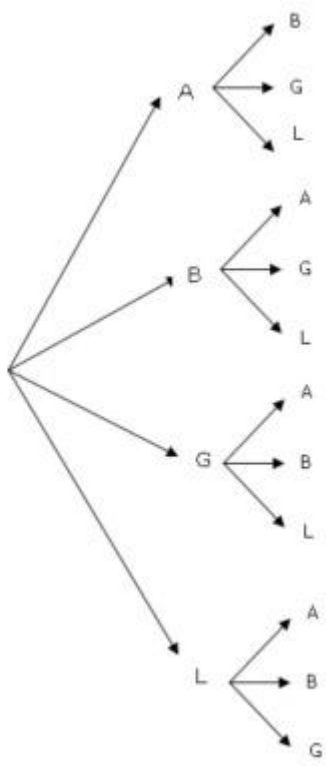
Situazione 1: Tra quattro studenti (Andrea, Barbara, Giorgio, Lucia) l'insegnante vuole scegliere il responsabile e il segretario del progetto fotografia. In quanti modi può fare tale scelta?

Situazione 2: Andrea deve sistemare 10 libri diversi ma ha solo 4 spazi liberi nella sua libreria. In quanti modi diversi può riempire a caso i 4 spazi?

Situazione 3: Quanti numeri di 4 cifre si possono scrivere usando le cifre **1, 2, 3, 4, 5**, **6** senza ripetere nessuna cifra?

Nella prima situazione un gruppo di studenti potrebbe decidere che il responsabile può essere scelto tra i quattro studenti mentre il segretario tra i restanti tre studenti, osservando che scegliere AB è diverso da BA avendo A e B nelle due scelte compiti diversi.

Un diagramma ad albero può essere usato come illustrazione visiva del perché vi sono 4 x 3 = 12 modi di scegliere i candidati.



I gruppo che risolve il secondo quesito potrebbe utilizzare un altro modello che potrebbe essere il seguente:

	1° spazio	2° spazio	3° spazio	4° spazio
Elementi fra cui scegliere	10	9	8	7

Nel primo spazio Andrea può mettere un libro a scelta tra 10, nel secondo uno tra i 9 rimasti, nel terzo spazio può metterne uno tra gli otto rimanenti ed, infine, nel quarto spazio un libro tra i 7 rimasti. Il totale dei modi diversi di scegliere è: 10 x 9 x 8 x 7 = 5040.

Anche il terzo problema può essere risolto con uno dei modelli precedenti ed in questo caso il totale dei modi diversi di scegliere sono è: 6 x 5 x 4 x 3 = 360

L'insegnante fa notare che nei tre problemi si voleva trovare il *numero delle scelte ordinate* di k elementi tra n oggetti tra loro diversi.

	1ª scelta	2ª scelta	3ª scelta	……	kª scelta
n. elementi fra cui scegliere			n – 2		– (k – 1)

Questo numero è dato dal prodotto di k fattori del tipo: n x (n – 1) x (n – 2) x ….x (n – (k – 1)).

Chiama queste situazioni **disposizioni semplici** e usa il simbolo D n,k che rappresenta il numero dei possibili modi per *scegliere* k elementi ordinati tra *n* oggetti tra loro diversi.

Quindi: $D_{n,k}$ = : n x (n – 1) x (n – 2) x …(n – (k – 1)).

L'insegnante fa notare che k <= n.

Fase 3

Permutazioni

L'insegnante propone alla classe, divisa in gruppi, una nuova serie di situazioni problematiche:

Situazione 1: In quanti modi Angelo, Carlo e Dario possono essere assegnati a giocare difensore, centrocampista o attaccante in una partita di calcio, se l'assegnazione è casuale?

Situazione 2: Tra quattro studenti (Andrea, Barbara, Giorgio, Lucia) l'insegnante deve scegliere il responsabile, il segretario, il relatore e il fotografo del progetto fotografia. In quanti modi possibili si può fare a caso tale scelta?

Un gruppo potrebbe risolvere la Situazione 1 usando una tabella in cui elencare tutte le possibili sistemazioni dei tre giocatori:

Difensore	Centrocampista	Attaccante	Ordinamenti possibili
Angelo	Carlo Dario	Dario Carlo	ACD ADC
Carlo	Angelo Dario	Dario Angelo	CAD CDA

Dario	Angelo Carlo	Carlo Angelo	DAC DCA

Un altro gruppo potrebbe risolvere lo stesso problema rappresentando le tre situazioni in questo modo dove le scelte possibili per il ruolo di difensore sono 3: A, C, D.

Difensore	Centrocampista	Attaccante
3		

Dopo aver scelto un giocatore per il ruolo di difensore, ne rimangono 2 fra i quali scegliere per giocare centrocampista. Per ciascuna scelta per il primo ruolo ne abbiamo due per il secondo.

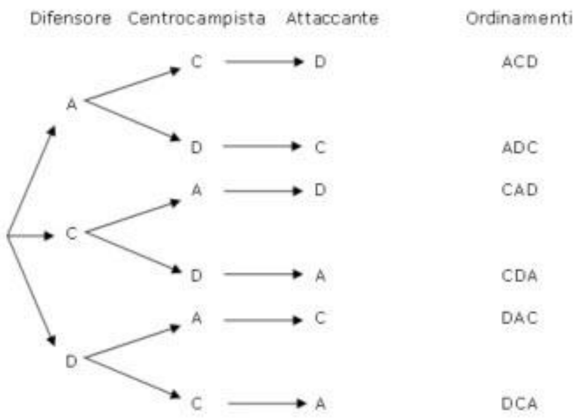
Difensore	Centrocampista	Attaccante
3	2	

Dunque si possono assegnare 3 giocatori al primo ruolo e 2 al secondo ruolo, dopo di che resta solo un giocatore da mettere nel terzo ruolo.

Difensore	Centrocampista	Attaccante
3	2	1

Il numero totale dei modi di assegnare i giocatori nei tre ruoli a caso è 3 x 2 x 1, cioè 6.

Un terzo gruppo potrebbe risolvere il problema usando il modello dell'albero:



Gli studenti osserveranno che oltre all'ordine, il numero degli elementi che formano la squadra (k) è uguale al numero degli studenti tra cui si sceglie di assegnare il ruolo (n). Gli studenti osservano che anche la Situazione 2 si risolve in modo analogo.

L'insegnante dice alla classe che nelle due situazioni le scelte si chiamano **permutazioni** ed il loro numero è indicato con Pn ed è pari a n! cioè: $n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots\ldots \cdot 3 \cdot 2 \cdot 1$

Fase 4

Combinazioni semplici

Continuando con la stessa metodologia, l'insegnante propone alla classe una nuova serie di problemi:

Situazione 1: Tra quattro studenti (Andrea, Barbara, Giorgio, Lucia) l'insegnante deve scegliere due ragazzi per giocare un doppio a tennis. In quanti modi possibili può fare tale scelta?

Situazione 2: In un compito in classe con quattro quesiti viene richiesto di risolverne tre. Quante scelte può fare uno studente scegliendo a caso i tre quesiti a cui rispondere?

Un gruppo potrebbe risolvere la prima situazione, osservando che, ad esempio, scegliere la coppia (Andrea, Lucia) o la coppia (Lucia, Andrea) è del tutto equivalente, e quindi pensando che, dalle 12 sequenze che erano emerse, ragionando come nella fase 3,si devono togliere 6 coppie, “doppioni delle precedenti” in quanto si tratta di coppie con gli stessi elementi che differiscono solo per l’ordine.

Pertanto le coppie risultanti sono: $\frac{4 x 3}{2 x 1} = \frac{12}{2} = 6$ coppie. Il che equivale a dire che delle possibili (precedenti) 12 scelte, la metà non devono essere considerate in quanto equivalenti alle precedenti.

Il gruppo che risolve la seconda situazione osserva che anche in questo caso l'ordine in cui vengono scelti i 3 esercizi, su 4, non è importante ed il numero delle scelte casuali possibili è 4. Come schema risolutivo l'insegnante invita gli studenti a scrivere in una tabella tutte le scelte ordinate di tre elementi fra 4; dove, per semplicità i quesiti del compito sono indicati con i numeri: 1, 2, 3, 4.

	Elementi	Scelte ordinate di 3 elementi
	1, 2, 3, 4	123, 213, 231, 132, 312, 321 124, 214, 241, 142, 412, 421 134, 314, 341, 143, 413, 431 234, 324, 342, 243, 423, 432
Numero	4	24

L'insegnante guida gli studenti ad osservare che in ogni riga, nella terza colonna della tabella, sono presenti le stesse scelte dei quesiti sia pure in diverso ordine il cui numero rappresenta le permutazione di 3 elementi: P_3. e vale 3! = 3 x 2 x 1:

Poiché l'ordine della scelta in cui vengono svolti gli esercizi non è importante basta prendere una sola terna por ogni riga e il numero di scelte *non ordinate* è 4.

L'insegnante chiama questo tipo di scelta: **combinazione semplice, il loro numero è indicato con Cn,k;** e ,nel caso analizzato, il numero delle scelte non ordinate è

$$\frac{\textit{numero delle scelte ordinate di 3 elementi su 4}}{\textit{numero delle permutazioni di 3 elementi}} = \frac{D_{4,3}}{P_3} = \frac{4 \cdot 3 \cdot 2}{3!} = \frac{24}{6} = 4$$

Generalizzando:

$C_{n,k} = \binom{n}{k} = \frac{n \cdot (n-1) \cdot (n-2) \cdots (n-(k-1))}{k!}$

Nota per insegnante:

L'insegnante può rivedere il problema da un altro punto di vista:

Nella tabella delle scelte ordinate di tre elementi su quattro definiamo la seguente relazione: "una scelta è equivalente ad un'altra se è formata dagli stessi elementi, sia pure in ordine inverso".

1. *la precedente relazione è una relazione di equivalenza in quanto gode delle proprietà: Riflessiva, Simmetrica e Transitiva;*
2. *le classi di equivalenza sono quelle elencate precedentemente e corrispondono alle scelte di tre elementi su quattro che sono indipendenti dall'ordine;*
3. *le classi di equivalenza sono quattro;*

Chiedersi in quanti modi è possibile scegliere k oggetti tra n, senza che interessi l'ordine di ogni scelta, equivale a chiedersi quanti sottoinsiemi di k elementi possiede un insieme di n elementi. Questo numero è indicato con il simbolo $\binom{n}{k}$.

Fase 5

Disposizioni con ripetizione

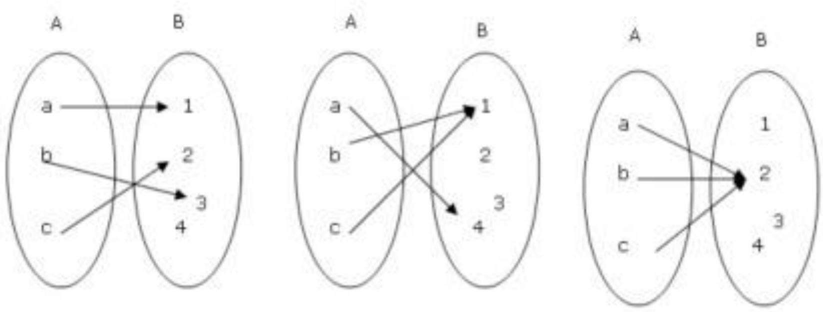
L'insegnante propone alla classe, divisa in gruppi, nuove situazioni problematiche.

Situazione 1: tra quattro studenti (Andrea, Barbara, Giorgio, Lucia) l'insegnante di matematica e fisica deve scegliere chi interrogare, in due giorni consecutivi, uno studente per giorno, mettendo i quattro nomi ogni volta in un sacchetto della tombola ed estraendone, a caso, uno. Quante diverse interrogazioni sono possibili?

Situazione 2: Siano dati i seguenti due insiemi: $A=\{a,b,c\}\ B=\{1,2,3,4\}$

Considera tutte le funzioni f: A → B, ossia considera le relazioni che associano a *ogni elemento* di A *uno e uno solo elemento* di B.

I diagrammi di Eulero seguenti sono esempi di tali funzioni:



Quante sono le funzioni che possiamo trovare?

Un gruppo fa notare che, nel primo problema, le diverse "interrogazioni" sono il prodotto cartesiano dei quattro elementi in quanto anche le coppie della diagonale principale, ad esempio la coppia (Barbara; Barbara), possono essere accettate perché ciò significa, ad esempio, che la "fortunata" Barbara viene interrogata due volte di seguito. Seguendo il ragionamento il numero delle coppie è in questa situazione 4 x 4= 16.

Un altro gruppo che risolve il secondo problema, dovendo contare quante sono le funzioni, può osservare che ogni elemento del primo insieme può avere come corrispondente un elemento qualunque del secondo insieme, e quindi in totale si hanno 4 x 4 x 4 = 4^3 = 64 *funzioni da A in B*.

L'insegnante chiama tali scelte **disposizioni con ripetizione** di n elementi presi a k a k. Il loro numero è la potenza di base n ed esponente k.

$D'_{n,k} = n^k$

L'insegnante fa notare che, adesso, k può essere >= n.

Fase 6

Permutazioni con elementi ripetuti

L'insegnante propone alla classe questi problemi:

Situazione 1 : in quanti modi è possibile anagrammare (anche ottenendo parole senza significato) la parola roma? E se la parola fosse bologna? E se fosse ferrara?

Dalla discussione gli studenti osservano che la parola roma presenta lettere tutte diverse e che i diversi anagrammi varieranno solo per l'ordine in cui le 4 lettere compaiono, pertanto si usano le permutazioni.
Da "roma" si possono individuare: 4! = 24 anagrammi.

Gli studenti che si occupano della parola "bologna" individuano che in quella vi sono due " o " e che fra gli anagrammi ve ne saranno alcuni ripetuti e cioè quelli che derivano dallo scambio delle due lettere o.

L'insegnante chiede quanti anagrammi si possono formare considerando le "o" diverse?

La risposta degli studenti è 7!

Quanti anagrammi se si scambiando le sole "o" fra loro? La risposta degli studenti è 2!

L'insegnante guida gli studenti alla soluzione: da "bologna" si possono individuare $\frac{7!}{2!} = \frac{5040}{2} = 2520$ anagrammi.

Un terzo gruppo di studenti osserva che la parola “ferrara” presenta tre “r” e due “a” e seguendo quindi il ragionamento fatto per la parola “bologna” rispondono che da “ferrara” il numero degli anagrammi possibili è $\frac{7!}{2!x3!}=\frac{5040}{2x6}=420$ anagrammi.
L'insegnante chiama questo tipo di associazione **permutazione con elementi ripetuti**.

Fase 7

Collegamenti
L’insegnante ricorda alla classe i diversi tipi di campionamento utilizzati nell’attività iniziale e chiede se è possibile individuare nelle tecniche di campionamento utilizzate per la costruzione dello spazio campionario di quelle attività alcune tra le nozioni di calcolo combinatorio sviluppate.
Gli studenti evidenziano che:

- nella Attività 1 Fase 2, per la costruzione dello spazio campionario, i cui dati sono riportati in tabella 3, in cui sia l’ordine sia la ripetizione sono necessarie, sono presenti le caratteristiche delle **disposizioni con ripetizione.**
- In **Spunti per un approfondimento disciplinare** nella attività 1 la tecnica usata per la costruzione dello spazio campionario, i cui dati sono riportati in tabella 14, presenta le caratteristiche delle **disposizioni semplici;**
- In **Spunti per un approfondimento disciplinare** nella attività 2 la tecnica usata per la costruzione dello spazio campionario, i cui dati sono riportati in tabella 18, richiama le **combinazioni semplici;**

Consulta l’approfondimento sul calcolo combinatorio

Nota per insegnante:

Il Calcolo Combinatorio dunque ci insegna a enumerare oggetti senza vederli ma tenendo in opportuna considerazione la modalità di scelta degli oggetti. Le situazioni precedenti , -- se parliamo di Calcolo Combinatorio --, non sono comunque esaustive, come dimostrano i seguenti problemi:

Problema 1 : Le possibili dimensioni dei lati di un triangolo sono: 6 cm, 7 cm, 8 cm, 9 cm. In quanti modi è possibile costruire un triangolo anche usando per i lati la stessa dimensione più volte?

Dalla discussione gli studenti osservano che si possono costruire 4 triangoli equilateri (scelta della dimensione del lato fra 4 possibili)

4x 3 =12 triangoli isosceli (dopo aver scelto la dimensione dei due lati uguali, in 4 modi, il terzo lato ha 3 possibili scelte)

4 triangoli scaleni (perché la scelta dei lati è indipendente dall'ordine e quindi equivale alla scelta di 3 elementi fra 4 nel senso delle combinazioni) in tutto 4+12+4= 20 triangoli.

Questo tipo di associazione, che prende in considerazione le "ripetizioni" e non prende in considerazione l'ordine, assume il nome di ***Combinazioni con ripetizioni*** *calcolabile anche con la formula* $C^r_{n,k} = C_{n+k-1,k}$

Abbiamo visto che il Calcolo Combinatorio è una scorciatoia per non fare tutti i "possibili conti": spesso però i conti bisogna proprio farli come nel seguente problema:

Problema 2 : In quanti modi diversi si possono collocare due Torri su una scacchiera?

Elementi per prove di verifica

Esercizio 1

Consideriamo una popolazione formata dalle quattro segretarie di un ufficio, ciascuna delle quali ha effettuato in una determinata giornata il numero di telefonate riportate nella tabella:

Segretaria	Numero telefonate effettuate
Anna	4
Carla	2
Daniela	7
Emma	5

Se si vogliono estrarre campioni casuali di numerosità 2 con reimmissione dalla popolazione delle segretarie, considerando l'ordine di estrazione:

a) Quanti sono i campioni possibili?

b) Come è composto l'universo dei campioni casuali possibili?

c) Calcolare per ogni campione il numero medio di telefonate.

d) Costruire la distribuzione di frequenze del numero medio di telefonate effettuate da un campione di due segretarie, la media e lo s.q.m. di tale distribuzione.

e) Confrontare la media della distribuzione del numero medio di telefonate effettuate da un campione di due segretarie con il numero medio di telefonate effettuate dalle quattro segretarie.

f) Se non conoscessimo il numero medio delle telefonate effettuate da ciascuna dalle quattro segretarie della popolazione, potremmo dire che la media delle medie di tutti i campioni coincide con la media (incognita) della popolazione?

Esercizio 2

Si vuole dare una “stima” al peso dei componenti di una classe composta da 26 studenti avendo a disposizione solo il peso di 6 studenti. Estratto a caso, con reinserimento, il seguente campione di pesi:

46, 50, 52, 57, 60, 65 Kg.

a) La sequenza proposta è l’unico campione che si può estrarre, con reinserimento dalla popolazione dei pesi della classe?

b) Si può ritenere che la media campionaria dei sei pesi ($\bar{x}_i$) sia uguale al peso medio (ignoto) della classe?

c) Qual è l'errore che si commette se si utilizza la media campionaria della sequenza dei pesi, sapendo che lo scarto quadratico medio dei pesi della classe è 6,83 kg?

d) Sapendo che lo scarto quadratico medio dei pesi della classe è 6,83 kg, quale dovrebbe essere la numerosità del campione se si desiderasse che la distribuzione della media campionaria dei pesi avesse una deviazione standard minore di 1,8 kg?

Esercizio 3

Un percorso pedonale tra Piazzale Roma e Rialto nella città di Venezia prevede l'attraversamento di sei ponti di cui sono stati contati i gradini:

Ponte	n. gradini
A	44
B	32
C	28
D	40
E	36
F	36

a) Scegliendo due ponti a caso fra i sei da percorrere, senza scegliere lo stesso ponte due volte, individuare lo spazio dei campioni possibile.

b) Per ogni campione individuare il numero massimo dei gradini dei ponti che lo compongono.

c) Il numero massimo di gradini in un campione (M_i) è il valore di una distribuzione campionaria?

d) Fornire la distribuzione di frequenze della variabile (M) numero massimo di gradini.

Esercizio 4

Una scatola contiene sei dischetti indistinguibili di cui tre hanno impresso su una faccia il valore 1, due hanno impresso il valore 2 ed uno il valore 0,5.

Scegliendo a caso due dischetti senza guardare, una alla volta con reinserimento:

a) Quanti e quali sono i campioni possibili?

b) Per ogni campione calcolare il valor medio.

c) Qual è la probabilità di aver estratto il campione (0,50 ;1)?

d) Qual è la probabilità di aver estratto un campione la cui media è 1,50?

e) Costruire la distribuzione dalla variabile casuale valore medio degli esiti ottenuti dal campionamento.

Esercizio 5

Si consideri un mazzo composto da 40 carte. Si estraggono simultaneamente 5 carte in modo da ottenere una mano di 5 carte. Quante sono le possibili mani? (Scegliere fra una delle risposte fornite).

a) $C_{40,5}$

b) 10^5

c) $D_{40,5}$

d) 5^{40}

Esercizio 6

Sette città sono collegate a due a due con una strada. Indicare il numero di strade, scegliendo fra una delle quattro risposte fornite.

a) 21

b) 42

c) 15

d) 49

Esercizio 7

Quante partite di scacchi singole possono essere giocate da cinque giocatori?

Esercizio 8

Quanti diversi "equipaggi" di una barca che richiede 3 ruoli diversi si possono formare, scegliendo a caso tra sette persone?

Esercizio 9

Sei persone hanno a disposizione sei sedie "in fila": in quanti modi diversi le possono occupare?

Soluzioni degli esercizi

Esercizio 1

a) I campioni possibili sono 4^2; (che rappresentano il numero delle disposizioni con ripetizione di 4 elementi presi a 2 a 2).

b)

	Anna	Carla	Daniela	Emma
Anna	(A;A)	(A;C)	(A;D)	(A;E)
Carla	(C;A)	(C;C)	(C;D)	(C;E)
Daniela	(D;A)	(D;C)	(D;D)	(D;E)
Emma	(E;A)	(E;C)	(E;D)	(E;E)

c)

	Anna	Carla	Daniela	Emma
Anna	4	3	5,5	4,5
Carla	3	2	4,5	3,5
Daniela	5,5	4,5	7	6
Emma	4,5	3,5	6	5

d)

$\bar{x}$	ni
2	1
3	2
3,5	2
4	1
4,5	4
5	1
5,5	2
6	2

7	1
Media	4,5 (telefonate)
s.q.m.	1,27 (telefonate)

a) I due valori coincidono e sono uguali a 4,5 telefonate.
b) Sì, perché la distribuzione delle medie campionarie fa riferimento all'intero spazio dei campioni.

Esercizio 2

a) No, in quanto il campione estratto è uno 26^6 elementi dello spazio dei campioni.
b) No, perché è il valor medio della sequenza campionaria estratta e, generalmente, $\bar{x}_i \neq \mu$.

c) Noto il valore della devst. della popolazione pari a 6,83 la valutazione dell'errore è data da $\frac{6{,}83}{\sqrt{6}} = 2{,}6$ kg.

d) Dalla relazione $\frac{\sigma}{\sqrt{n}} < 1{,}8$, noto □, si ricava n > 14.

a)

	A	B	C	D	E	F
A		(A;B)	(A;C)	(A;D)	(A;E)	(A;F)

B	(B;A)		(B;C)	(B;D)	(B;E)	(B;F)
C	(C;A)	(C;B)		(C;D)	(C;E)	(C;F)
D	(D;A)	(D;B)	(D;C)		(D;E)	(D;F)
E	(E;A)	(E;B)	(E;C)	(E;D)		(E;F)
F	(F;A)	(F;B)	(F;C)	(F;D)	(F;E)	

b)

	A	B	C	D	E	F
A		44	44	44	44	44
B	44		32	40	36	36
C	44	32		40	36	36
D	44	40	40		40	40
E	44	36	36	40		36
F	44	36	36	40	36	

c) Sì, in quanto al variare del campione varia a caso il massimo trovato, che dipende dal campione casuale considerato.

d)

Distribuzione del massimo campionario	
Mi	ni
32	2
36	10
40	8
44	10
Totale	30

Esercizio 3

a)

	A	B	C	D	E	F
A		(A;B)	(A;C)	(A;D)	(A;E)	(A;F)
B	(B;A)		(B;C)	(B;D)	(B;E)	(B;F)
C	(C;A)	(C;B)		(C;D)	(C;E)	(C;F)
D	(D;A)	(D;B)	(D;C)		(D;E)	(D;F)
E	(E;A)	(E;B)	(E;C)	(E;D)		(E;F)
F	(F;A)	(F;B)	(F;C)	(F;D)	(F;E)	

b)

	A	B	C	D	E	F
A		44	44	44	44	44
B	44		32	40	36	36
C	44	32		40	36	36
D	44	40	40		40	40
E	44	36	36	40		36
F	44	36	36	40	36	

c) Sì, in quanto al variare del campione varia a caso il massimo trovato, che dipende dal campione casuale considerato.

d)

Distribuzione del massimo campionario	
Mi	ni
32	2
36	10
40	8
44	10
Totale	30

Esercizio 4

a) Considerando i dischetti diversi solo per il valore riportato sulla faccia, il numero dei campioni possibili è 3^2 ed ogni coppia ha diversa probabilità di essere estratta che dipende dal numero di dischetti per ciascun valore.

	0,5	1	2
0,5	(0,5;0,5)	(0,5;1)	(0,5;2)
1	(1; 0,5)	(1;1)	(1;2)
2	(2; 0,5)	(2;1)	(2;2)

b)

	0,5	1	2
0,5	0,5	0,75	1,25
1	0,75	1	1,5
2	1,25	1,5	2

c) Poiché il dischetto con impresso il valore 0,5 è unico mentre i dischetti con valore 1 sono tre la probabilità di estrarre la coppia (0,5; 1) è $\frac{1}{6} \cdot \frac{3}{6} = \frac{3}{36}$

d) Dalla tabella del punto b si osserva che $\bar{x}$ = 1,5 si ottiene per le coppie ordinate (1;2) e (2;1). Quindi la probabilità chiesta è: $\frac{3}{6} \cdot \frac{2}{6} + \frac{2}{6} \cdot \frac{3}{6} = \frac{12}{36}$

e) La distribuzione delle medie campionarie è una distribuzione di probabilità come da tabella che segue.

Distribuzione delle medie campionarie	
$\bar{x}$	$p(\bar{x})$
0,5	$\frac{1}{36}$
0,75	$\frac{6}{36}$
1	$\frac{9}{36}$
1,25	$\frac{4}{36}$
1,5	$\frac{12}{36}$
2	$\frac{4}{36}$

Esercizio 5

a) $C_{40,5}$ In quanto non interessa l’ordine in cui le carte sono assegnate al giocatore.

b) 10^5

c) $D_{40,5}$

d) 5^{40}

Esercizio 6

a) 21 Per collegare a due a due sette città si possono usare le combinazioni e il numero delle strade è dato da $C_{7,2} = \frac{D_{7,2}}{P_2} = \frac{7 \cdot 6}{2} = 21$

b) 42

c) 15

d) 49

Esercizio 7

Si deve valutare il numero $C_{5,2}$ delle combinazioni semplici di 5 elementi di classe 2

(presi 2 a 2: una partita a scacchi viene infatti giocata da due giocatori):

$$C_{5,2} = \frac{D_{5,2}}{P_2} = \frac{5 \cdot 4}{2} = 10$$

Esercizio 8

Essendo 3 i ruoli diversi da ricoprire, il numero di scelte casuali possibili è dato dalle disposizioni di 7 elementi presi a tre a tre. $D_{7,3} = 7 \cdot 6 \cdot 5 = 210$

Esercizio 9

Sono le permutazioni di sei elementi $P_6 = 6! = 720$.

Nota: se le sedie fossero disposte "in tondo" le possibilità sarebbero $P_5 = 5! = 120$.

Risorse

Bibliografia

Ferrari, P., Nicolini, G., Tommasi, C. *ntroduzione all'inferenza statistica*. G. Giappichelli Editore, Torino 2006.

Frosini, B.V., Montinaro, M., Nicolini, G. *Il campionamento da popolazioni finite*. UTET libreria, Torino 1994.

Orsi, R. *Probabilità e inferenza statistica*. Il Mulino, Bologna 1985.

Gnednko, B. *Teoria della probabilità*. Editori Riuniti *Univ. Press,* Roma 2011.

Baldi, P. *Calcolo delle probabilità e statistica.* McGraw-Hill Companies, 2003.

Cicchitelli, G. *Probabilità e Statistica*. Maggioli editore, Santarcangelo di Romagna 2004.

Boggio, A., Borello, G., *Statistica*, vol. 1 e 2, Petrini editore, 2008.

Quinn, R.J., Wiest, L.R. *Una linea costruttiva nell'insegnare permutazioni e combinazioni*. In: Induzioni, 41, 2010, pp.137-146.

Matematica 2004 *La matematica del cittadino* (Campioni e forchette - Piccoli campioni crescono).

Sitografia

Vichi, M. Campionamento e attendibilità dei risultati di un'indagine
(Visitato nel giugno 2013)

Istat: Metodologie e tecniche per la produzione delle stime di interesse e la valutazione degli errori campionari
(Visitato nel giugno 2013)

Linee guida metodologiche per le rilevazioni statistiche
(Visitato nel giugno 2013)