Campionamento e Distribuzioni Campionarie

Introduzione

- Introdurre le indagini campionarie
- Analizzare il le tecniche di costruzione dei campioni e di rilevazione
- Sviluppare il concetto di distribuzione campionaria
- Sviluppare il concetto di media campionaria

Alcuni elementi di base

Popolazione	↔	Campione
↓		↓
Parametri		Statistiche campionarie
μ σ^2		$\bar{X}$ S^2
N		n

Formazione del campione - 1

Estrazione con reintroduzione (ripetizione)

è un tipo di estrazione in cui ogni unità statistica estratta viene rilevata e reintrodotta nella popolazione, e, dunque, può essere nuovamente estratta in un momento successivo

Estrazione senza reintroduzione (ripetizione)

ogni unità statistica estratta viene rilevata e non viene più reintrodotta nella popolazione, e, quindi, non può essere nuovamente estratta in un momento successivo.

Formazione del campione - 2

Campione ordinato

è un campione che differisce da un altro semplicemente se cambia la posizione delle unità statistiche all'interno del campione: quindi, il campione AB è differente dal campione BA

Campione non ordinato

è un campione che differisce da un altro se cambia almeno una delle unità statistiche all'interno del campione: quindi, ad esempio, il campione AB è differente dal campione dal campione AC, mentre è uguale al campione BA.

Formazione del campione - 3

Campione casuale

un campione estratto in modo casuale è un campione nel quale ogni unità statistica ha una determinata probabilità di essere estratta

È l'estrazione casuale che permette di far aderire ognuno dei campioni potenzialmente estraibili da una popolazione ad una legge di probabilità

Campionamento casuale semplice

Immaginare di avere una sorta di grande urna all'interno della quale ci sono tante palline.

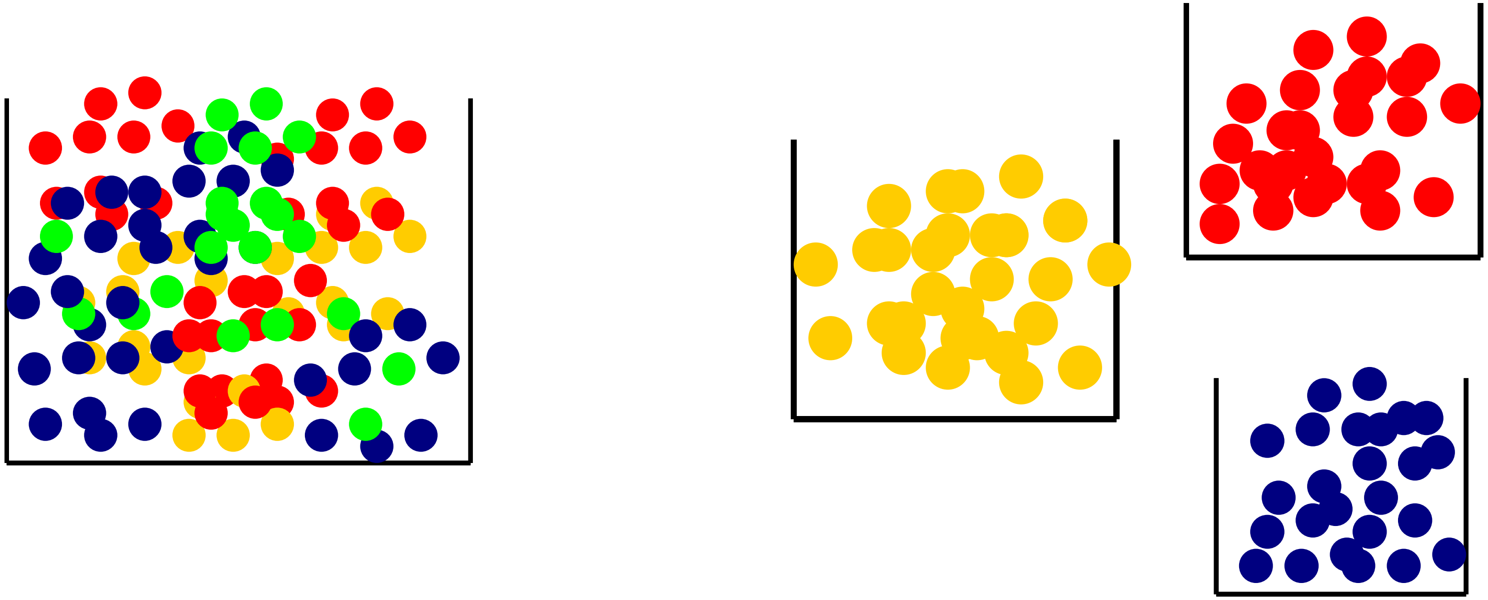
L'elemento fondamentale è che, ad ogni estrazione, ogni pallina ha la stessa probabilità di essere estratta dall'urna e che, in generale, ogni campione che è possibile formare (della medesima dimensione, ovviamente) ha la stessa probabilità di essere estratto

Abbiamo bisogno:

1) **conoscere tutte le unità statistiche che compongono la popolazione**
2) **tali unità devono essere tutte reperibili**
3) **estrarre casualmente le unità**

Campionamento casuale stratificato

Si tratta di suddividere la popolazione di riferimento in gruppi (gli "strati") sulla base della (eventuale) conoscenza del fenomeno investigato, in modo tale da "costruire" dei gruppi omogenei a seconda di determinate caratteristiche



Campionamento casuale stratificato - 2

Quali sono i vantaggi della stratificazione? Costruendo gruppi "omogenei" al loro interno, ed estraendo da tali gruppi le unità statistiche sarà possibile:

1) **ottenere stime non solo di tutta la popolazione, ma anche dei "sottogruppi"**
2) **migliorare le stime, nel senso di renderle più "precise"**
3) **in "alternativa" al punto due, a parità di precisione delle stime, sarà possibile ottenere la stessa "informazione" a partire da una numerosità campionaria inferiore (e, quindi, risparmiamo!)**

La logica dell’inferenza

Popolazione ⟷ Campione

↑ ↓

Generale ← Particolare

Ma ci sono tanti modi per estrarre un campione di numerosità n

Ognuno ci da un risultato

Ma allora è solo fortuna?

Certo che no! Bisogna scoprire la “regolarità” (o la regola)

Generale → Particolare

Universo dei campioni

Supponiamo di avere a disposizione una popolazione di 5 pazienti del nostro ospedale dei quali conosciamo la pressione sistolica, che riportiamo qui di seguito :

A = 140 ; B = 160 ; C = 120 ; D = 180 ; E = 150

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{(140+160+120+180+150)}{5} = 150$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n} = \frac{(140-150)^2 + (160-150)^2 + (120-150)^2 + (180-150)^2 + (150-150)^2}{5} = 400 \qquad \sigma = 20$$

UNIVERSITÁ DEGLI STUDI DI TERAMO

Universo dei campioni - 2

A = 120 ; B = 180 ; C = 150 ; D = 160 ; E = 140

N^n

n=2 — **campione ordinato estratto con reintroduzione**

		x_1	x_2	$\bar{x}$
A	A	120	120	120
A	B	120	180	150
A	C	120	150	135
A	D	120	160	140
A	E	120	140	130
B	A	180	120	150
B	B	180	180	180
B	C	180	150	165
B	D	180	160	170
B	E	180	140	160
C	A	150	120	135
C	B	150	180	165
C	C	150	150	150

		x_1	x_2	$\bar{x}$
C	D	150	160	155
C	E	150	140	145
D	A	160	120	140
D	B	160	180	170
D	C	160	150	155
D	D	160	160	160
D	E	160	140	150
E	A	140	120	130
E	B	140	180	160
E	C	140	150	145
E	D	140	160	150
E	E	140	140	140

X_i	n_i	$P(X_i)$
120	1	0,04
130	2	0,08
135	2	0,08
140	3	0,12
145	2	0,08
150	5	0,20
155	2	0,08
160	3	0,12
165	2	0,08
170	2	0,08
180	1	0,04
	25	1,00

UNIVERSITÁ DEGLI STUDI DI TERAMO

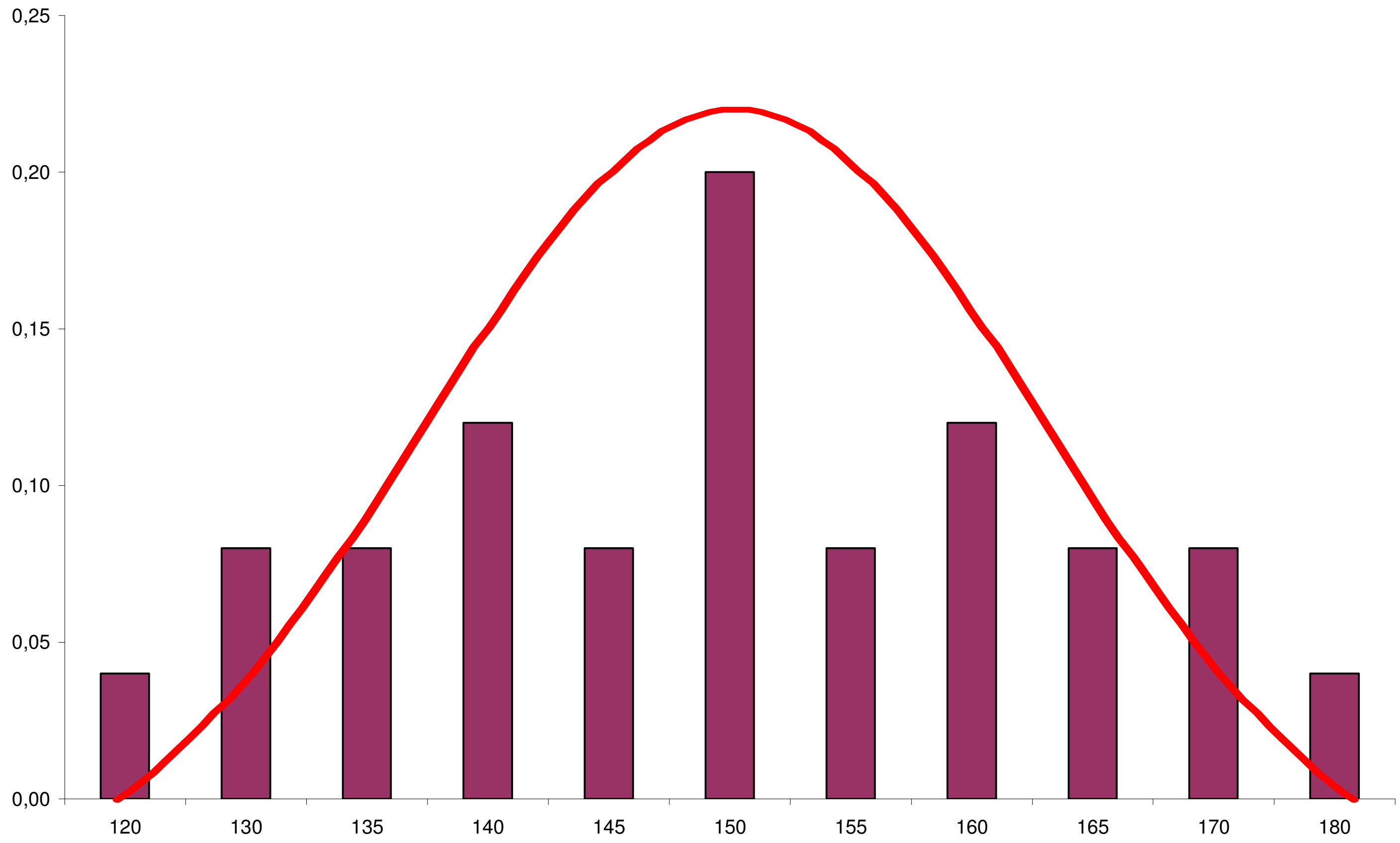
V.C. Media Campionaria

X_i	n_i	$P(X_i)$	X_i*$P(x_i)$	$[X_i-E(X_i)]^2$	$[X_i-E(X_i)]^2$ x $P(x_i)$
120	1	0,04	4,8	900	36
130	2	0,08	10,4	400	32
135	2	0,08	10,8	225	18
140	3	0,12	16,8	100	12
145	2	0,08	11,6	25	2
150	5	0,20	30,0	0	0
155	2	0,08	12,4	25	2
160	3	0,12	19,2	100	12
165	2	0,08	13,2	225	18
170	2	0,08	13,6	400	32
180	1	0,04	7,2	900	36
	25	1,00	150		200

$$\mathrm{E}\left(\overline{\mathrm{X}}\right) = 150$$

$$\mathrm{VAR}\left(\overline{\mathrm{X}}\right) = 200$$

V.C. Media Campionaria - Grafico



V.C. Media Campionaria

Variabile casuale Media Campionaria

è quella variabile casuale costruita prendendo in considerazione tutte le possibili medie che possiamo calcolare a partire da tutti i possibili campioni di una certa numerosità estraibili da una determinata popolazione.

Si distribuisce come la curva normale

Ha media pari alla media vera della popolazione

$$E(\overline{X}) = \mu$$

Ha varianza pari alla varianza "vera" della popolazione diviso la numerosità campionaria

$$VAR(\overline{X}) = \frac{\sigma^2}{n}$$

$$s.q.m.(\overline{X}) = \frac{\sigma}{\sqrt{n}}$$

Alcune ulteriori considerazioni - 1

Che succede se estraiamo il campione in modo differente?

Nel caso di estrazione senza reintroduzione, e campione non ordinato:

Distribuzione Normale

$$E(\overline{X}) = \mu$$

$$\left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n}$$

estrazione con reintroduzione e campione ordinato

$$\frac{\sigma^2}{n}$$

Variabilità minore!!

Alcune ulteriori considerazioni - 2

Posso utilizzare sempre la normale? O dipende da come si distribuiscono i dati di partenza?

qualsiasi sia la distribuzione di partenza, la distribuzione della media campionaria tende alla distribuzione normale all'aumentare della dimensione campionaria.

Teorema del LIMITE CENTRALE

Riferimenti sul testo

di **Whitlock M.C., Schluter D.**
***Analisi statistica dei dati biologici*, Zanichelli**

Paragrafi da studiare: 10.5.
Esercizi alla fine dei paragrafi.