

DISTRIBUZIONE CAMPIONARIA

Una distribuzione campionaria è la distribuzione dei valori di una statistica campionaria.

Per esempio, si abbia la seguente tabella di dati distribuiti su una popolazione:

x	n
1	12
2	34
3	40
4	45

Calcoliamo la media (*indicata da μ perché si riferisce a popolazione*):

$$\mu = \frac{\sum xn}{\sum x} = \frac{380}{131} = 2,9$$

Ora estraiamo, per esempio, 5 valori a caso e facciamone la media (*indicata da \bar{x} soprassegnato perché riferita a campione*):

I valori potrebbero essere: 1 1 3 4 2 (*guardando la tabella in cui ci sono 12 “1”, 34 “2” ecc.*)
la cui media è:

$$\bar{x} = \mathbf{2,2}$$

Chiamiamo “**stimatore**” il parametro “media” di questo campione, definendolo in genere una funzione delle variabili aleatorie X_1, X_2 ect:

$$T=f(X_1, X_2, ..., X_n)$$

e chiamiamo “**stima**” il suo valore (2,2 nel nostro caso)

- 1^a osservazione: ogni estrazione campionaria avrà valori diversi e valori diversi di stima.
- 2^a osservazione: di solito il parametro μ è ignoto.

DISTRIBUZIONE CAMPIONARIA DELLE MEDIE

Esempio:

La popolazione è costituita dai 6 valori ottenibili dal lancio di un dado.

1 2 3 4 5 6

I suoi parametri sono: la media μ e la varianza σ^2

Valore caratteristico di
dispersione

$\mu = (1+2+3+4+5+6)/6 = 3,5$ media della popolazione

[ricordiamo questo valore]

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N} = [(1-3,5)^2 + (2-3,5)^2 + \dots] / 6 = 2,916$$

Tutti i possibili campioni di ampiezza $n=2$ sono $D'_{N,n} = N^n = 6^2 = 36$, ovvero le facce:

1 1
1 2
1 3
1 4... Ecc.

Numeriamo le facce in modo da poter fare la media di ciascuno dei campioni:

$(1+1)/2 = 1$
 $(1+2)/2 = 1,5$
 $(1+3)/2 = 2$
 $(1+4)/2 = 2,5$... Ecc.

Otteniamo così la seguente

DISTRIBUZIONE CAMPIONARIA DELLE MEDIE

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

Possiamo calcolare la media di questa distribuzione campionaria di medie:

$$m_{\bar{x}} = \frac{\sum \bar{x} f}{\sum f} = \frac{(1*1) + (1,5*2) + (2*3) + \dots + (6*1)}{1+2+3+4+5+6+5+4+3+2+1} = 3,5$$

[Media delle medie]

Possiamo calcolare la varianza:

$$\sigma_{\bar{x}}^2 = \frac{\sum (\bar{x} - \mu_x)^2 f}{\sum f} = 1,458$$

E notiamo che il parametro della popolazione $\mu=3,5$ è uguale alla media della distribuzione campionaria $\mu_{\bar{x}}=3,5$

Ciò vale anche in generale, e pertanto:

La media della distribuzione campionaria delle medie coincide con la media della popolazione.

$$\mu_{\bar{x}} = \mu$$

Esempio 2:

6 individui hanno i seguenti pesi: 46, 50, 52, 57, 60, 65 Kg.

La media vale $\mu=55$

La varianza è

$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n} = \frac{(46-55)^2 + (50-55)^2 + (52-55)^2 + (57-55)^2 + (60-55)^2 + (65-55)^2}{6} = 40,67 \text{ varianza della}$$

popolazione

Otteniamo dalla popolazione di origine tutti i campioni di dimensione 2 estraendo con reimmissione tutte le possibili coppie, che sono $D'_{N,n} = N^n = 6^2 = 36$, vale a dire:

46,46	46,50	46,52	46,57	46,60	46,65
50,46	50,50	50,52	50,57	50,60	50,65
52,46	52,50	52,52	52,57	52,60	52,65
57,46	57,50	57,52	57,57	57,60	57,65
60,46	60,50	60,52	60,57	60,60	60,65
65,46	65,50	65,52	65,57	65,60	65,65

Calcoliamo le medie \bar{x} di ciascuno dei 36 campioni:

46	48	49	51,5	53	55,5
48	50	51	53,5	55	57,5
49	51	52	54,5	56	58,5
51,5	53,5	54,5	57	58,5	61
53	55	56	58,5	60	65
55,5	57,5	58,5	61	62,5	65

Questa è una distribuzione campionaria di medie.

Se ne potevano ottenere altre se avessimo estratto campioni di dimensione 3, 4... Perciò diremo che questa è “una” delle distribuzioni possibili.

$$\text{La media della distribuzione delle medie è } \mu_{\bar{x}} = \frac{\sum \bar{x}}{n} = \frac{1982}{36} = 55$$

$$\text{La varianza della distribuzione delle medie è } \sigma_{\bar{x}}^2 = \frac{\sum (\bar{x} - \mu_{\bar{x}})^2}{n} = \frac{775}{36} = 21,54$$

Dividiamo la varianza della popolazione **40,67** per la varianza della distribuzione di medie **21,54**:

$$\sigma^2 / \sigma_{\bar{x}}^2 = 1,89 \text{ (assimilabile a 2)}$$

Ricordando che $n=2$ (abbiamo detto che tutti i campioni hanno dimensione 2)
Si può verificare che

$$\sigma^2/\sigma_x^2 = n \text{ [in realtà era 1,86 assimilabile a 2]}$$

da cui con opportuni passaggi

$$\frac{\sigma^2}{\sigma_x^2} = n \text{ ricavo}$$
$$\sigma^2 = n\sigma_x^2$$

e

$$\sigma_x^2 = \sigma^2/n$$

da cui, estraendo la radice, la distribuzione campionaria delle medie risulta essere

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

Il parametro $\frac{\sigma}{\sqrt{n}}$ viene chiamato **errore tipico o errore standard**, dove σ è la deviazione standard della popolazione di origine e n è l'effettivo dei campioni estratti.

Inoltre:

Al crescere del numero degli elementi del campione, la distribuzione campionaria delle medie assume una deviazione standard sempre minore.

Esempio: calcolare media e deviazione standard dei numeri 10,20,40,60,100 e media e dev.st. della distribuz. camp. delle medie con campioni di $n = 2$; $n=5$; $n=50$.

$$\mu = 46$$

$$\sigma^2 = 5120/5 = 1024$$

$$\sigma = 32$$

Come detto prima, la media della distr.camp. delle medie e la media della popolazione risultano uguali

$$\mu_x = \mu = 46$$

La deviazione standard della distr.camp. delle medie dipende da n :

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

$$\text{con } n=2 \quad 32/1,41=22,7$$

$$\text{con } n=5 \quad 32/2,24=14,28$$

$$\text{con } n=50 \quad 32/7,07=4,53$$

Si nota che

al crescere di n (che così si avvicina a N della popolazione di origine), tanto più i valori della media \bar{x} del campione si avvicinano alla media μ della popolazione (essi coincideranno quando $n=N$) e la deviazione standard $\sigma_{\bar{x}}$ diminuisce.

Ciò significa che le distribuzioni delle medie campionarie sono tanto più concentrate attorno a μ quanto più n è grande: **quanto più la dimensione del campione è grande, tanto più è probabile che la media del campione sia simile alla media della popolazione.**

Ciò ci fa esporre il Teorema del Limite centrale così:

La distribuzione di medie campionarie \bar{x} di campioni con dimensioni n tende alla

Distribuzione Normale $N(\mu, \frac{\sigma}{\sqrt{n}})$ quando n è sufficientemente grande (in genere $n \geq 30$).

Ricordando che la variabile Z (normale standard) corrispondente al Teorema del Limite centrale

[in una somma di var. al. indipendenti S con medesima distr. di prob., media e varianza, al crescere di n detta somma S tende a $N(n\mu, n\sigma^2)$ si legge normale con media $n\mu$, e varianza $n\sigma^2$]

assume il valore

$$Z = \frac{S - n\mu}{\sqrt{n}\sigma}$$

e dal momento che $S = X_1 + X_2 + \dots + X_n$, per cui $\frac{S}{n} = \frac{X_1 + X_2 + \dots + X_n}{n} = \bar{X}$ Quindi $S = n\bar{X}$

E allora, sostituendo

$$Z = \frac{n\bar{X} - n\mu}{\sqrt{n}\sigma} \text{ e dividendo numeratore e denominatore per } n$$

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Se la popolazione di origine ha una distribuzione con media μ e deviazione standard σ , la media campionaria \bar{X} avrà distribuzione normale (con n grande) con media μ_x e deviazione standard σ_x

dove

$$\mu_x = \mu$$

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

Esempio:

L'oggetto A ha una durata di 230 ore con deviazione di 36 ore. Se si esaminano a caso 40 oggetti, determinare la probabilità che la loro durata media sia <233 ore.

$$\mu = 230$$

$$\sigma = 36$$

$$\bar{x} = 233$$

$$n = 40$$

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = Z = \frac{233 - 230}{\frac{36}{\sqrt{40}}} = 0,53$$

Consultando la Tavola della funzione di ripartizione della normale ridotta, si ottiene:

$$P(\bar{X} \leq 233) = P(Z \leq 0,53) = 0,7019$$

ovvero, in percentuale, $P=70,19\%$