Statistica inferenziale

STATISTICA NUMERICA

A.Y. 2022-2023

Outline

1)Campionamento (sampling)

2)Distribuzioni campionarie
e
teorema del limite centrale

1)Campionamento (sampling)

Analisi statistica di dati

Nella analisi statistica dei dati si utilizzano dei campioni di dati per inferire (dedurre) delle informazioni sulla popolazione da cui il campione (i campioni) sono stati estratti.

Popolazione: tutti gli elementi di un data set

Campione: una o piu osservazioni relative alla popolazione.
Il campione viene scelto casualmente (Simple Random Sample)

Analisi statistica di dati

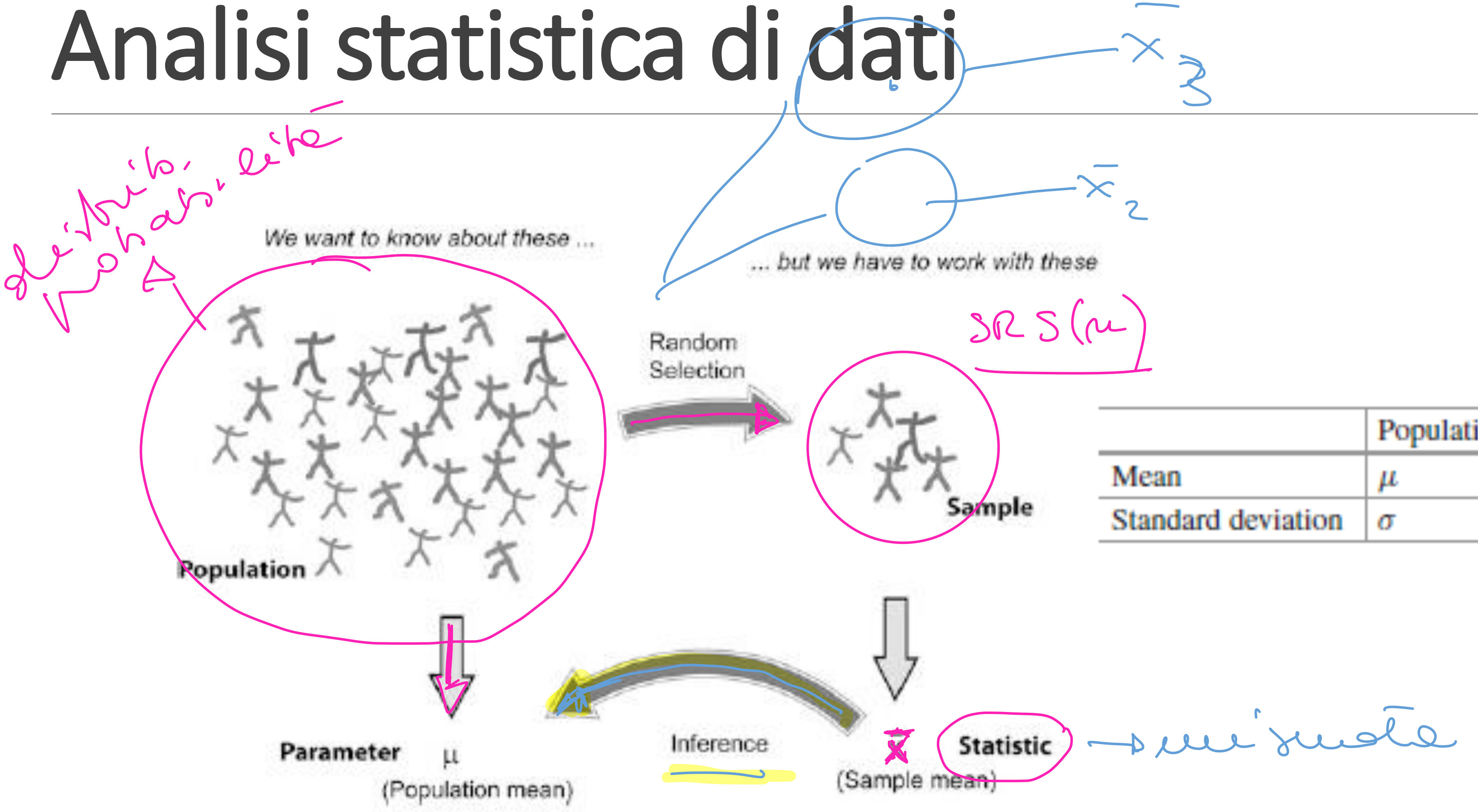
Parametro: valore caratteristico di una popolazione, come la media, la deviazione standard.
(di solito indicati con lettere greche)

Statistica: valori misurabili delle caratteristiche di un campione, come la media, la deviazione standard il massimo e il minimo.

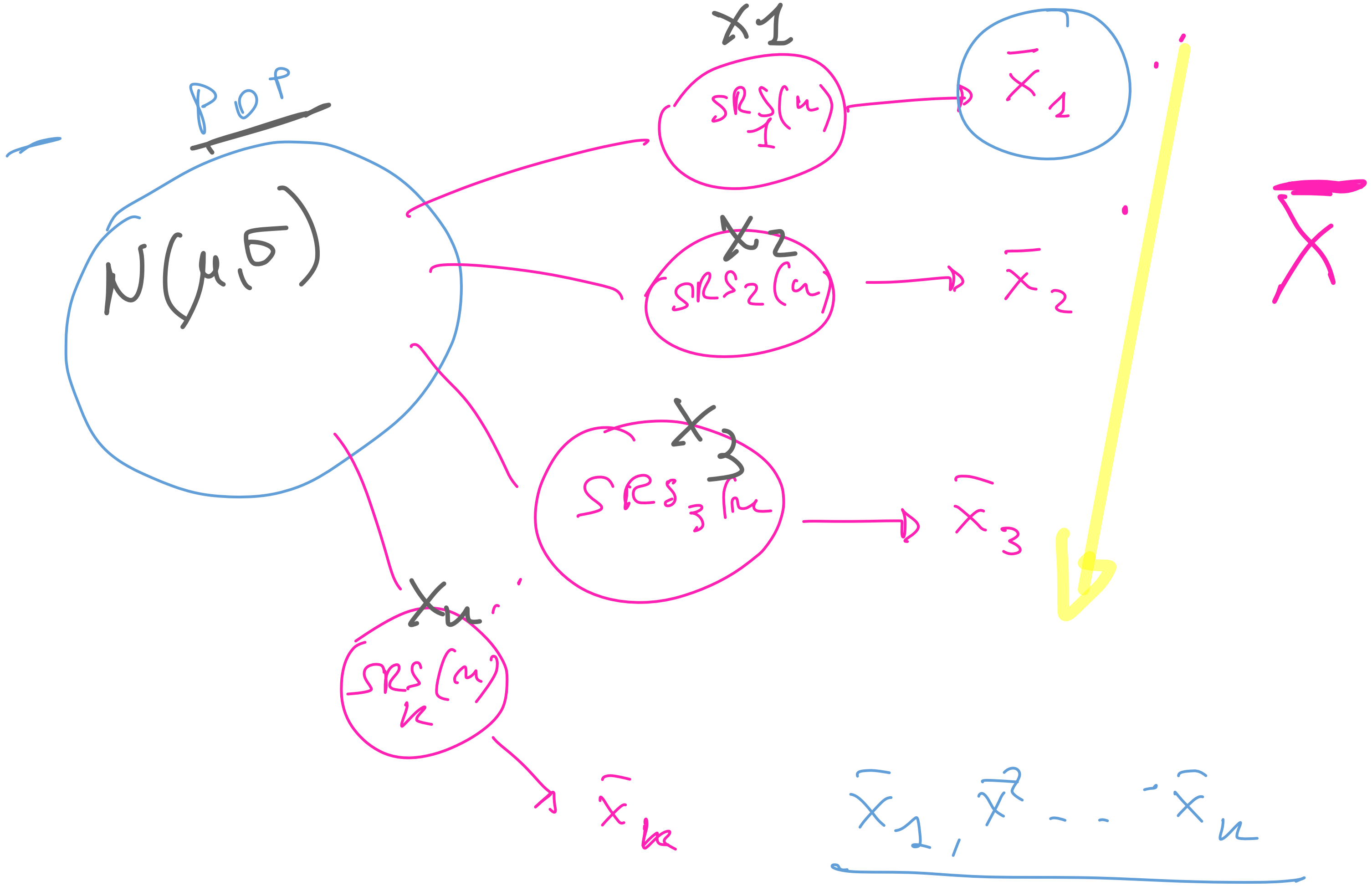
Distribuzione dei campioni: La distribuzione di una statistica (misurata sui campioni)

Statistica inferenziale: stimare uno o piu parametri della popolazione utilizzando la statistica dei campioni

Analisi statistica di dati



	Population parameter	Sample statistic
Mean	μ	$\bar{x}$
Standard deviation	σ	s


pop
$N(\mu, \sigma)$
X_1
$SRS_1(n)$
$\bar{x}_1$
X_2
$SRS_2(n)$
$\bar{x}_2$
X_3
$SRS_3(n)$
$\bar{x}_3$
X_n
$SRS_k(n)$
$\bar{x}_k$
$\bar{X}$
$\bar{x}_1, \bar{x}_2 \ldots \bar{x}_n$

Sampling

Il campione scelto deve essere:

1. casuale- scelto in modo random dalla popolazione
2. Rappresentativo – deve coprire le i diversi valori delle caratteristiche considerate
3. Di dimensione adeguata- non troppo piccolo rispetto alla varianza dei valori considerati
4. Non bias - non ci devono essere distorsioni rispetto alla statistica da misurare.

2)Distribuzioni campionarie
e
teorema del limite centrale

CALCOLO MEDIA CAMPIONARIA

$$SRS(n) = x_1, x_2, \dots x_n$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

VARIANZA CAMPIONARIA

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

DEVIAZIONE STANDARD CAMPIONARIA

$$s = \sqrt{s^2}$$

$x_i = i \qquad x_1 = 1, \; x_2 = 2 \ldots \quad x_3 = 3 \ldots$

$SRS(7) = \{1, 2, 3, 4, 5, 6, 7\}$

$\bar{x} = \frac{28}{7} = 4$

$s^2 = \frac{1}{7}\left[(1-4)^2 + (2-4)^2 + \cdots\right] =$

$\frac{1}{7}\left(3^2 + 2^2 + 1^2 + 0^2 + 1^2 + 2^2 + 3^2\right)$

$s = \sqrt{s^2}$

Distribuzioni campionarie: media campionaria

Siano $X_1, X_{2, \dots} X_n$ SRS(n) da una distribuzione aleatoria NORMALE con (media=mu, sd=sigma). Allora la Allora:

1. la variabile aleatoria $\bar{X}$ media campionaria ha distribuzione normale (media=mu, sd=sigma/ $\sqrt{n}$).

2. La variabile aleatoria:

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

Ha distribuzione campionaria **normale standard.**

$\sim N(0,1)$

$N(\mu, \sigma)$

$\bar{X}_1 \to \bar{x}_1$

$\bar{X}_2 \to \bar{x}_2$

$\bar{X}_3 \to \bar{x}_3$

$\vdots$

$\bar{X}_n$

$\bar{X}_n$

$\bar{X}$ media campionaria

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Distribuzioni campionarie: varianza campionaria

normale

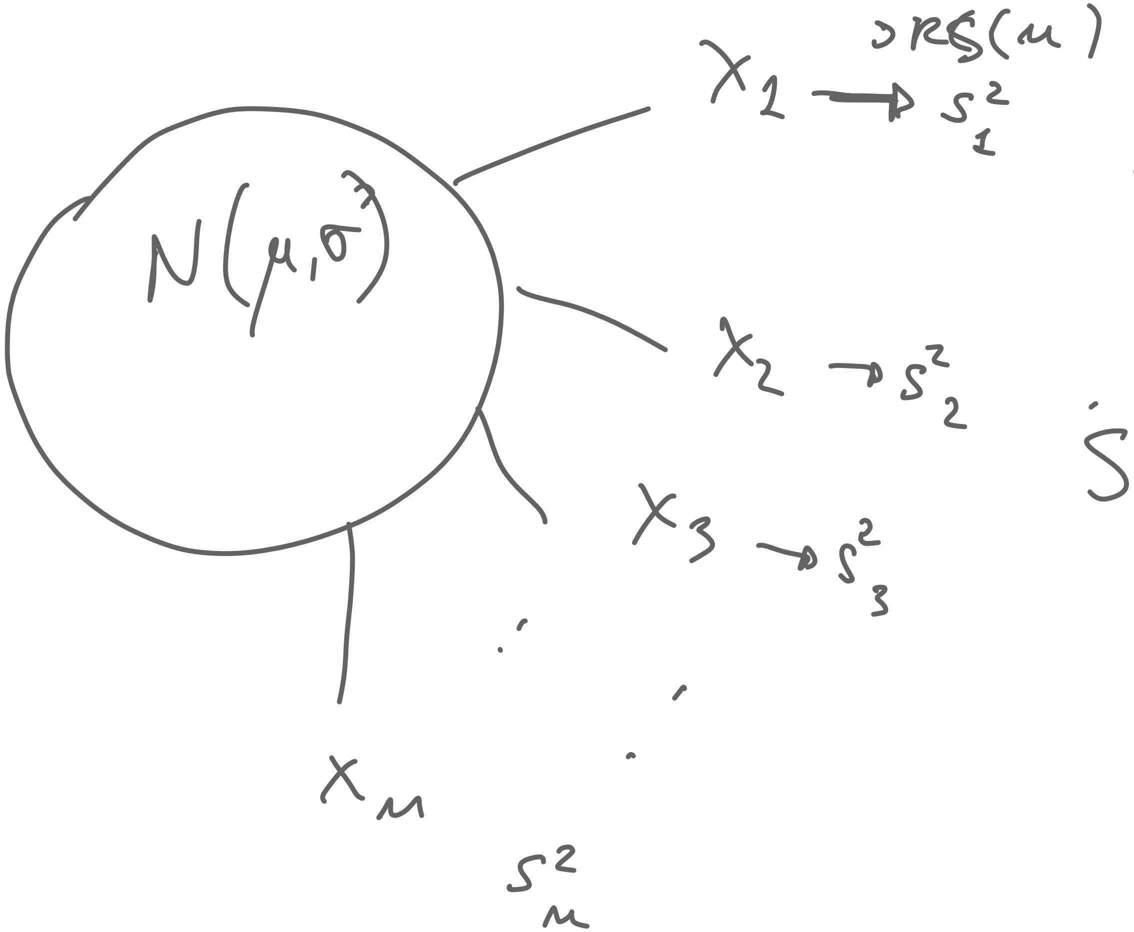
Siano $X_1, X_2, \dots X_n$ SRS(n) da una distribuzione aleatoria con (media=mu, sd=sigma). Allora la variabile aleatoria S^2 varianza campionaria

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

scalata:

$$\frac{n-1}{\sigma^2}S^2$$

Ha distribuzione chiquadro (df=n-1).



S^2 var. aleatoria

$S^2 \sim \chi^2(df = n-1)$

Teorema del limite centrale

QUALUNQUE

Teorema del limite centrale. Siano $X_1, X_2, \ldots X_n$ una SRS(n) da una distribuzione di popolazione con media μ e deviazione standard σ. Allora la variabile aleatoria

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

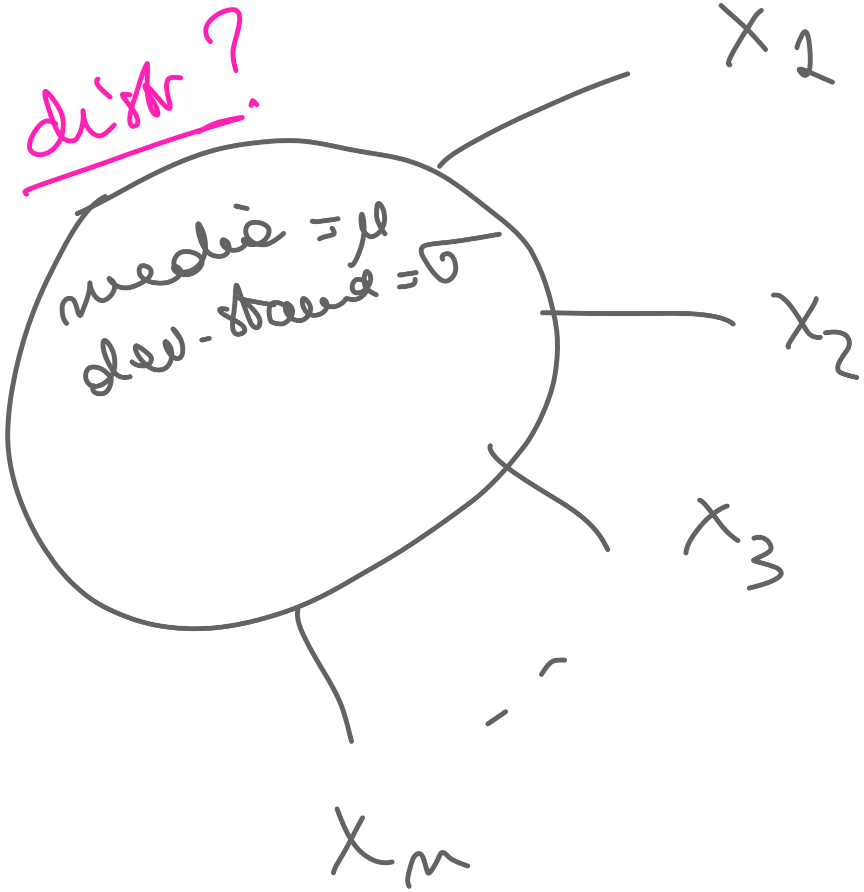
$\bar{x}$ var. media campionaria

ha una distribuzione campionaria che **ha come limite**, per $n \to \infty$, la distribuzione normale standard (`norm(mean = 0, sd = 1)`).

$$Z \underset{n \to \infty}{\sim} N(0,1)$$

equivalentemente

(2) $\bar{X} \underset{n \to \infty}{\sim} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$



$\bar{X} \underset{n \to \infty}{\sim} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

$n \geq 30-40$

Se anche non conosco la distribuzione della popolazione, se ho un campione sufficientemente grande posso supporre (per quanto riguarda il comportamento delle variabili aleatorie $\bar{X}$ e S^2) la distribuzione di popolazione come <u>NORMALE</u>.

Teorema del limite centrale

Teorema del limite centrale. Siano $X_1, X_2, \ldots X_n$ una SRS(n) da una distribuzione di popolazione con media μ e deviazione standard σ. Allora la variabile aleatoria

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

ha una distribuzione campionaria che **ha come limite**, per $n \to \infty$, la distribuzione normale standard (`norm(mean = 0, sd = 1)`).