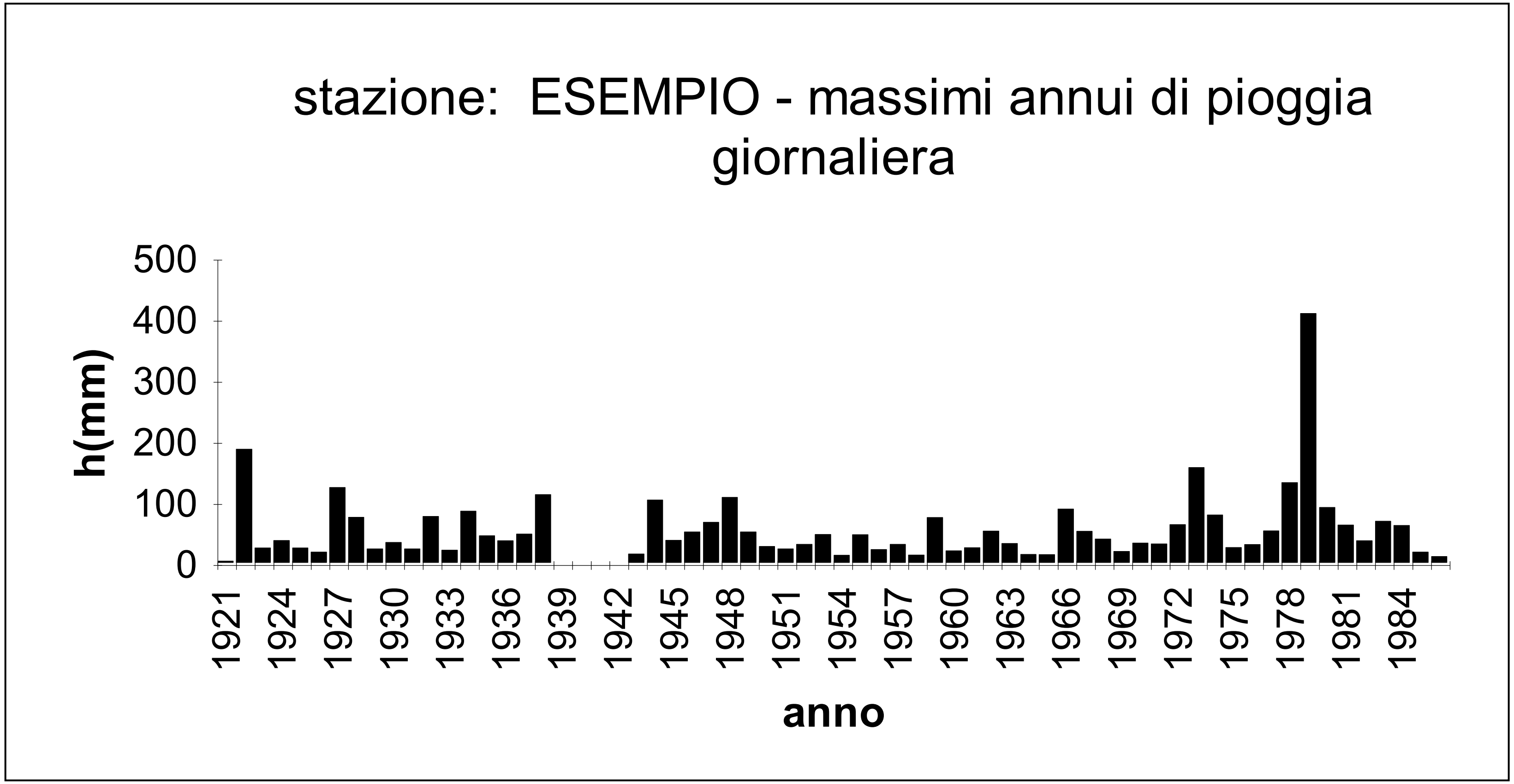
ANALISI ESPLORATIVA

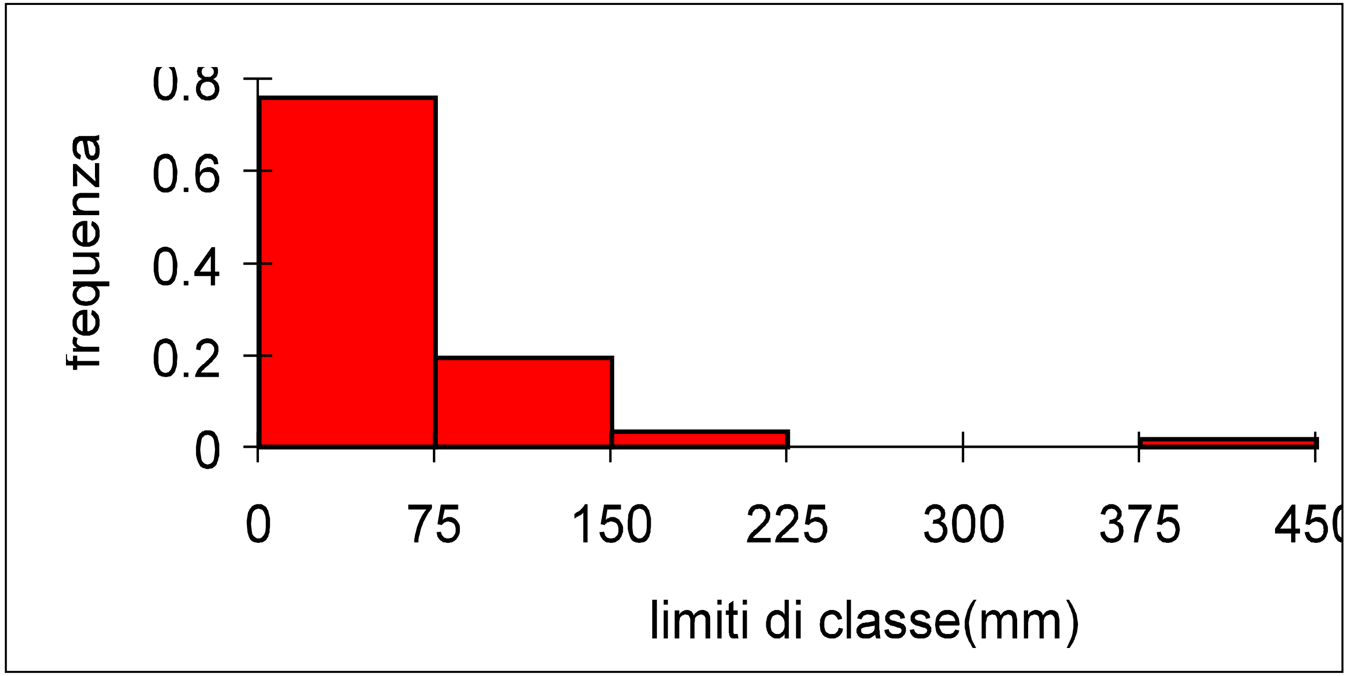
DELLE SERIE DI OSSERVAZIONI

Rappresentazione grafica della serie storica (successione cronologica)



Ampiezza del campione $A = (x_{max} - x_{min})$

e *rappresentazione ad istogramma* delle **frequenze di classe**, sia assolute che relative.



Onde evitare l'arbitrarietà della determinazione del numero di classi, si può utilizzare la relazione suggerita da *Sturges* che lega il numero delle classi, *k*, alla dimensione del campione, *n*, secondo la relazione: $k = \text{int}(1 + 3.3 \log N)$

(logaritmo espresso in base 10)

MOMENTI CAMPIONARI

Media campionaria

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

Varianza

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

Coefficiente di asimmetria (skewness)

$$Ca = \frac{\frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^3}{s^3}$$

Coefficiente di appiattimento (kurtosi)

$$\kappa = \frac{\frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^4}{s^4}$$

*vengono usati come **stime** dei momenti teorici della distribuzione da cui si ipotizza provengano le osservazioni*

Se t è la stima e θ è il valore vero del parametro (es. momento di ordine r), le due quantità si avvicinano se (N=numero osservaz):

- t è una stima consistente di θ quando:

$$|t-\theta| < \varepsilon \quad \forall\varepsilon \qquad \text{per} \qquad N \to \infty$$

- t è non distorta se:

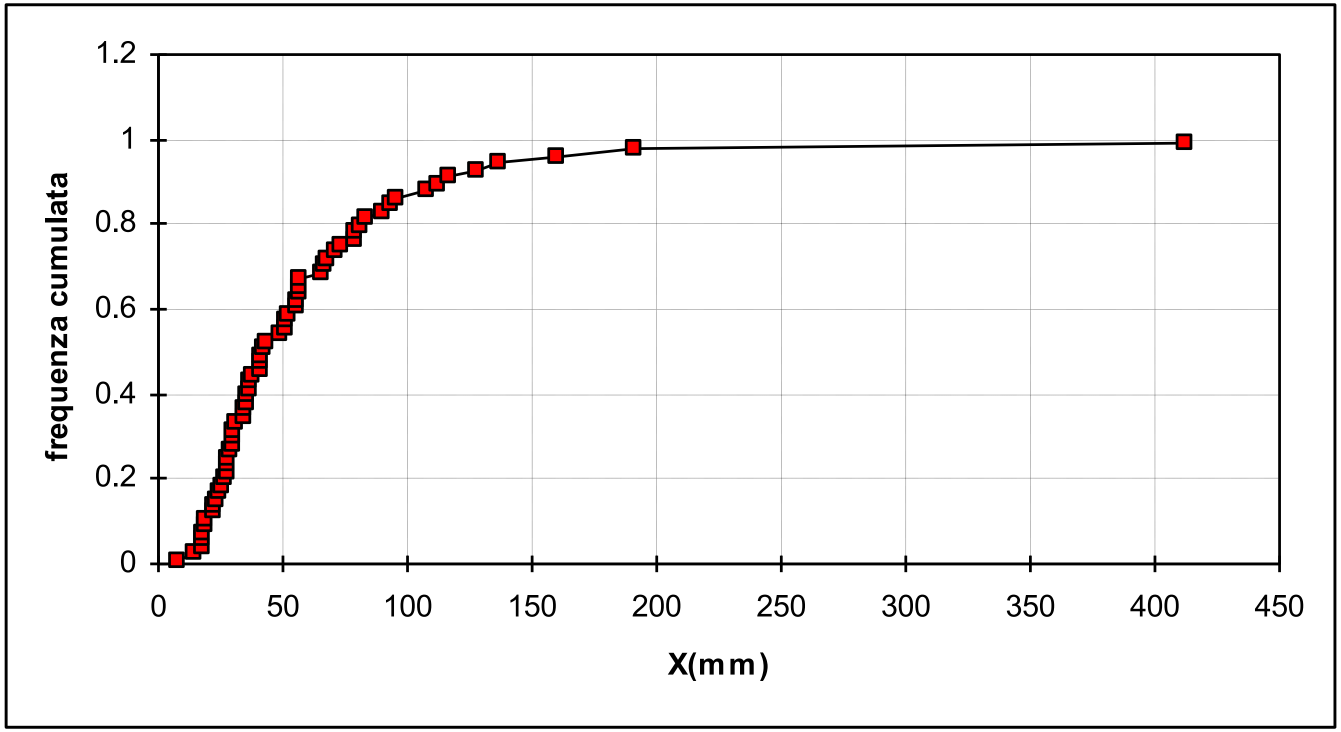
$E(t) = \theta \ \ \forall n$ (*asintoticamente*, se vale solo per $N \to \infty$)

<u>Correzioni</u>:

per varianza: $$\widehat{var} = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

per asimmetria $$\widehat{Ca} = \frac{(n-2)^2}{(n)(n-1)} \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\widehat{var}^{2/3}}$$

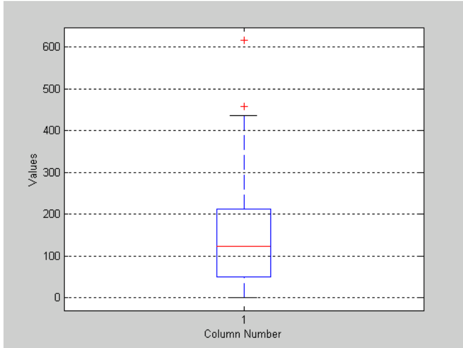
Curva di *Frequenza cumulata* delle osservazioni



Frequenza cumulata campionaria:

$$\phi(x_i) = \frac{i}{n}$$

Rappresentazione Box-Plot della serie



Limiti del *box*:

Inferiore: I quartile del campione $x(\Phi=0.25)$

Superiore: III quartile del campione $x(\Phi=0.75)$

Linea mediana: II quartile del campione $x(\Phi=0.50)$

Si definisce *range interquartile* (**IQR**) la differenza:

$$\boldsymbol{IQR} = X(F=0.75) - X(F=0.25)$$

Limiti dei *whiskers*:

Inferiore

Valor <u>minimo</u> della serie delle osservazioni (X_1)

oppure

I quartile - 1.5 volte ***IQR*** → $X(F=0.25)$ - 1.5 ***IQR***

Se negativo può essere posto pari a zero quando le osservazioni sono definite positive

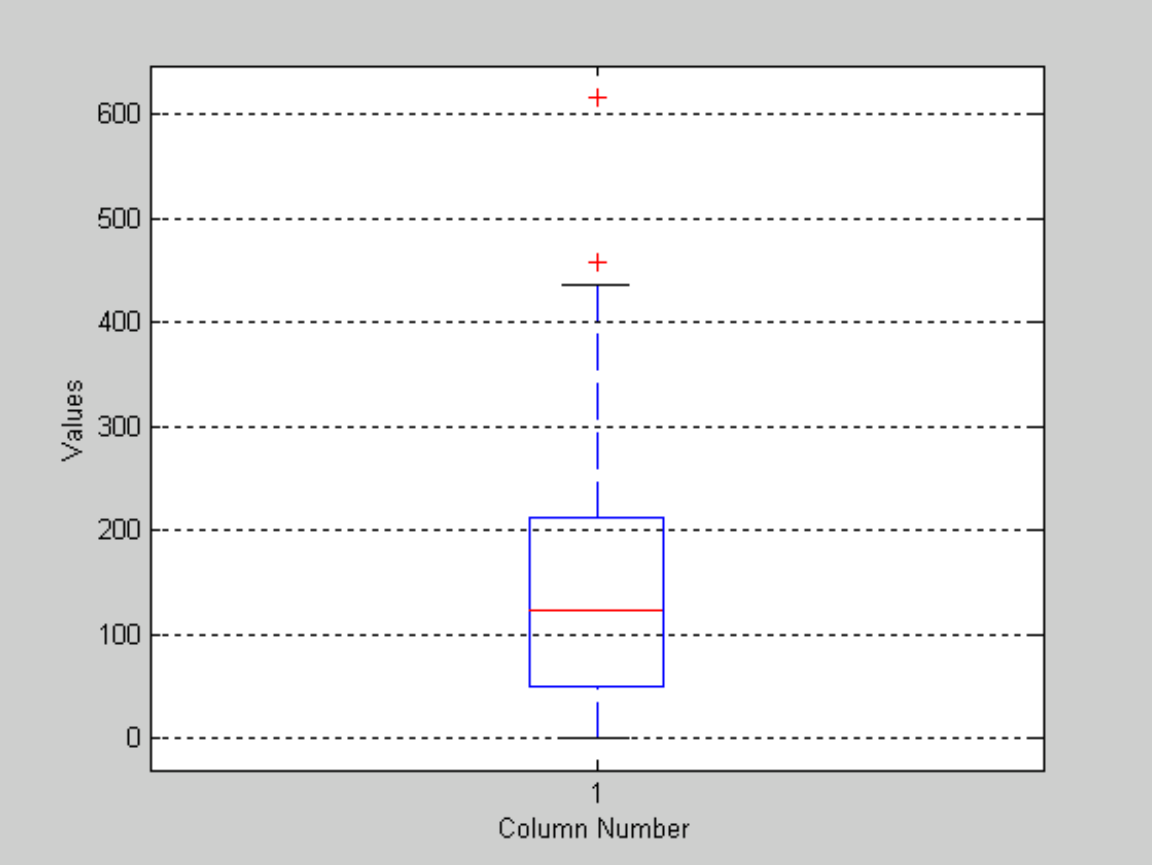
Superiore

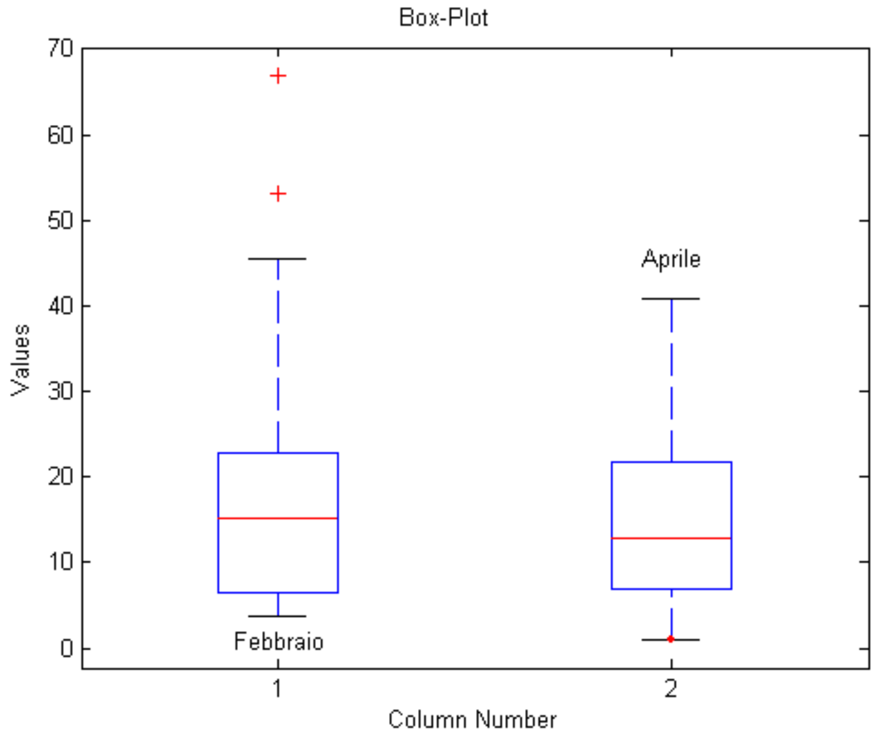
Valor <u>massimo</u> della serie delle osservazioni (X_n)

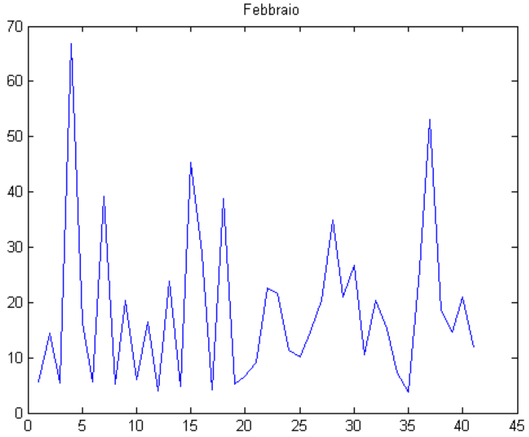
oppure

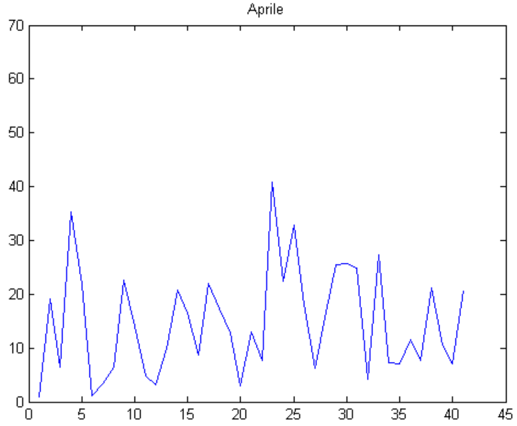
III quartile + 1.5 volte ***IQR*** → $X(F=0.75)$ + 1.5 ***IQR***

Nella rappresentazione con i *whiskers* si possono indicare tutte le osservazioni di valore inferiore al *whisker* minimo e superiore al *whisker* massimo




Box-Plot
Values
Column Number
Febbraio
Aprile
0
10
20
30
40
50
60
70
1
2


Febbraio


Aprile