Università degli Studi di Napoli «Federico II»
Facoltà di Architettura

U*p*ta

Corso di laurea in Urbanistica e Scienze della Pianificazione Territoriale e Ambientale

AA 2013-2014

Corso integrato di Matematica e statistica

Docente: Livia D'Apuzzo.

Statistica Descrittiva Univariata

4. *Distribuzioni di frequenza cumulata e funzione di ripartizione per caratteri misurabili in una scala ordinale (caratteri quantitativi e caratteri qualitativi ordinati)*

FREQUENZE CUMULATE E DISTRIBUZIONI DI FREQUENZE CUMULATE PER CARATTERI QUANTITATIVI

Sia il carattere oggetto di un'indagine statistica quantitativo e sia x un suo valore; ha senso chiedersi quante sono le unità statistiche in cui la determinazione del carattere non supera x (cioè risulti $\leq x$): ad esempio se il carattere è il reddito annuo, posto x = 25.000 euro, ha senso chiedersi quante famiglie hanno un reddito annuo minore o uguale a 25.000 euro.

Accanto alla frequenza assoluta del valore x di un carattere, definita da

$f_a(x) = n_x$ ***= numero delle unità con determinazione carattere uguale a*** x

= numero dei casi in cui il carattere assume valore uguale a x***,***

consideriamo la* frequenza (assoluta) cumulata al valore** x ***così definita:

$f_{cum}(x) = N_x$ ***= numero unità con determinazione carattere minore o uguale a*** x

= numero dei casi in cui il carattere assume valore minore o uguale a x***.***

Sia X l'insieme dei valori rilevati del carattere; la funzione che a ogni x in X associa la frequenza assoluta cumulata a x

$$x \in X \rightarrow f_{cum}(x).$$

è strettamente crescente: $x < y \Rightarrow f_{cum}(x) < f_{cum}(y)$.

Supponiamo che x e y siano due valori del carattere effettivamente rilevati e che sia $x < y = x+h$, $h>0$. Poiché ogni valore rilevato $t \leq x$ verifica la condizione $t < y$, i casi in cui il carattere assume valore minore o uguale a x sono compresi tra i casi in cui il carattere assume valore minore o uguale a y, ma sono in numero minore, quindi $f_{cum}(x) < f_{cum}(y)$. Ad esempio supponiamo che in una indagine sul reddito annuo di un certo numero di famiglie siano stati rilevati redditi di 25.000 euro e di 30.000 euro; le famiglie che hanno un reddito minore o uguale a 25.000 hanno anche un reddito minore di 30.000 euro, pertanto l'insieme A delle prime famiglie è incluso nell'insieme B delle famiglie con reddito minore o uguale a 30.000, ma non coincide con esso poiché ci sono famiglie con reddito rilevato di 30.000.

A	**Famiglie con reddito che non supera 25.000 euro**	$\subset$	**Famiglie con reddito che non supera 30.000 euro**	***B***

Allora $f_{cum}(25.000)$ ***= numerosità di A*** $< f_{cum}(30.00)$ ***= numerosità di B.***

E' possibile contare il numero di unità statistiche in cui la determinazione del carattere non supera il valore x, anche se x non è effettivamente un valore rilevato; si definisce pertanto $f_{cum}(x)$ *per ogni* x *in* $R =]-\infty, +\infty[$ e si considera *la funzione*

$$f_{cum}: x \in R \rightarrow f_{cum}(x) \qquad (1)$$

è detta **distribuzione semplice di frequenze assolute cumulate**

f_{cum} *è una funzione crescente :* $x < y \Rightarrow f_{cum}(x) \leq f_{cum}(y)$;

può quindi accadere che $x < y$ e $f_{cum}(x) = f_{cum}(y)$.[(*)]

Frequenze relative cumulate e funzione di ripartizione.
Sia n il numero totale delle unità statistiche costituenti la popolazione o il campione indagato. La **porzione** di popolazione o campione che presenta il carattere in una determinazione minore o uguale al valore x è ovviamente data dal rapporto tra $f_{cum}(x)$, frequenza cumulata a x, ed n

Il rapporto

$F(x) = f_{cum}(x)/n$ è detto ***frequenza relativa cumulata a x***

La funzione che rappresenta la distribuzione delle frequenze relative cumulate è

$$F: x \in R \rightarrow F(x) = f_{cum}(x)/n \qquad (2)$$

ed è detta **funzione di ripartizione**

Frequenze percentuali cumulate e corrispondente distribuzione.

La **percentuale** delle unità statistiche per le quali la determinazione del carattere ha valore minore o uguale a x si ottiene moltiplicando per 100 la frequenza relativa cumulata ed è detta ***frequenza percentuale cumulata*** *a x*

$F_{per}(x) = (f_{cum}(x)/n) \times 100$ ***frequenza percentuale cumulata a x***

La funzione che rappresenta la distribuzione delle frequenze percentuali cumulate è:

$$F_{per}: x \in R \rightarrow F_{per}(x) = (f_{cum}(x)/n) \times 100 = F(x) \times 100 \qquad (3)$$

ed è detta **distribuzione di frequenze percentuali cumulate**

(*) --------
Ad esempio se in una indagine sul reddito sono stati rilevati i redditi 23.000 euro e 25.000 euro e nessun reddito compreso tra essi è stato rilevato, allora il numero dei casi riscontrati con reddito che non supera 24.000 è uguale al numero dei casi con reddito che non supera 23.000, cioè $f_{cum}(24.000) = f_{cum}(23.000)$

3.1 - CALCOLO DELLE FREQUENZE CUMULATE LIMITATAMENTE AI VALORI RILEVATI PER IL CARATTERE QUANTITATIVO DISCRETO .

Consideriamo il caso che il carattere sia discreto e che assuma solo un numero finito di valori $x_1, x_2, \dots x_k$:

$X = \{x_1, x_2, \dots x_k\}$ **insieme dei valori assunti dal carattere**

Per calcolare le frequenze cumulate i valori devono essere indicati in ordine crescente; pertanto supponiamo che

$$x_1 < x_2 < \dots < x_k.$$

<u>Calcolo delle frequenze assolute cumulate a valori effettivamente rilevati</u>

Per definizione

$f_a(x_i) = n_i$

n_i =*numero delle unità con determinazione del carattere* $= x_i$

$f_{cum}(x_i) = N_i$

N_i =*numero delle unità con determinazione del carattere* $\le x_i$.

Iª regola di calcolo delle frequenze assolute cumulate a valori rilevati*: la frequenza cumulata a un valore* x_i *è somma della frequenza assoluta di* x_i *e delle frequenze dei valori che sono minori di* x_i*. Per determinare allora le frequenze cumulate, occorre ordinare i valori* $x_1, x_2, \dots x_n$ *del carattere in ordine crescente e utilizzare le frequenze assolute* $f_a(x_i)$

Supposto $x_1 < x_2 < \dots < x_i < \dots < x_n$ risulta:

x_i	$n_i = f_a(x_i)$	$f_{cum}(x_i)$
x_1	n_1	n_1
x_2	n_2	$n_1 + n_2$
$\vdots$	$\vdots$	
x_i	n_i	$n_1 + n_2 + \dots + n_{i-1} + n_i$
$\vdots$	$\vdots$	
x_k	n_k	$n_1 + n_2 + \dots + n_{k-1} + n_k = n$
	n	

$f_{cum}(x_1) = f_a(x_i) = n_1$

$f_{cum}(x_2) = f_a(x_1) + f_a(x_2) = n_1 + n_2$

....

$$f_{cum}(x_i) = f_a(x_1) + f_a(x_2) + \dots + f_a(x_{i-1}) + f_a(x_i) = n_1 + n_2 + \dots + n_{i-1} + n_i \qquad (4)$$

....

$f_{cum}(x_k) = f_a(x_1) + f_a(x_2) + \dots + f_a(x_k) = n_1 + n_2 + \dots + n_k = n$ *<u>frequenza totale</u>.*

E quindi

$$f_{cum} : x_i \in X = \{x_1, x_2, \dots x_k\} \rightarrow f_{cum}(x_i) = N_i = \sum_{h=1}^{i} n_h$$

Esempi

Distribuzione degli studenti iscritti al primo anno di una facoltà universitaria secondo l'età

x_i= età degli studenti	$f_a(x_i)$= numero studenti di età x_i	$f_{cum}(x_i)$= numero studenti di età $\leq x_i$ $f_{cum}(x_i) = f_a(x_1) + f_a(x_2) + \dots + f_a(x_{i-1}) + f_a(x_i)$
x_1=19	880	880
x_2=20 anni	1200	2080= 880+1200
x_3=21 anni	600	2680= 880+1200+600
x_4=22 anni	400	3080=880+1200+600+400
x_5=23 anni	320	3400=880+1200+600+400+320
	n= 3400	

Chilometri percorsi in areo in un anno

valore x_i Chilom. ↓	Freq.ass. $f_a(x_i)$	Freq. Cum. $f_{cum}(x_i)$
0	15	15
480	1	16=15+1
1250	1	17=15+1+1
4800	1	18=15+1+1+1
6600	1	19=15+1+1+1+1
9300	1	20=15+1+1+1+1+1
Tot.	*20*	

$$f_{cum}(0) = 15 = f_a(0)$$

$$f_{cum}(480) = 16 = 15+1 = f_a(0) + f_a(480)$$

$$f_{cum}(1250) = 17 = 15+1+1 = f_a(0) + f_a(480) + f_a(1250)$$

$$f_{cum}(4800) = 18 = 15+1+1+1 = f_a(0) + f_a(480) + f_a(1250) + f_a(4800)$$

$$f_{cum}(6600) = 19 = 15+1+1+1+1 = f_a(0) + f_a(480) + f_a(1250) + f_a(4800) + f_a(6600)$$

$$f_{cum}(9300) = 20 = f_a(0) + f_a(480) + f_a(1250) + f_a(4800) + f_a(6600) + f_a(9300)$$

UN DIVERSO MODO DI CALCOLARE LE FREQUENZE CUMULATE

Da $f_{cum}\ (x_i) = \underbrace{f_a\,(x_1)+ f_a\,(x_2)+\ldots+ f_a\,(x_{i-1})} + f_a\,(x_i)$ *otteniamo*

$$f_{cum}\ (x_i) = f_{cum}\,(x_{i-1}) + f_a\,(x_i) \quad \text{per } i>1$$

IIª regola di calcolo delle frequenze assolute cumulate: ***la frequenza assoluta cumulata a un valore x_i è somma della frequenza assoluta di x_i e della frequenza cumulata al valore x_{i-1} immediatamente precedente.***

Allora le uguaglianze ***(4)*** possono essere così riscritte

$$f_{cum}\ (x_1) = f_a\,(x_i) = n_1$$

$$f_{cum}\ (x_2) = f_{cum}\,(x_1)+ f_a\,(x_2)$$

....

$$f_{cum}\ (x_i) = f_{cum}\,(x_{i-1}) + f_a\,(x_i) \qquad (5)$$

....

$$f_{cum}\ (x_k) = f_{cum}\,(x_{k-1}) + f_a\,(x_k) = n \qquad \text{frequenza totale}$$

Distribuzione degli studenti iscritti al primo anno di una facoltà universitaria secondo l'età

x_i *= età degli studenti*	$f_a(x_i)$ *= numero studenti di età* x_i	$f_{cum}(x_i)$ *= numero studenti di età* $\leq x_i$ $f_{cum}(x_i) = f_{cum}\,(x_{i-1}) + f_a\,(x_i)$ *per* $i>1$
x_1*=19*	*880*	*880*
x_2*=20 anni*	*1200*	*2080= 880+1200*
x_3*=21 anni*	*600*	*2680= 2080+600*
x_4*=22 anni*	*400*	*3080=2680+400*
x_5*=23 anni*	*320*	*3400=3080+320*
	n*= 3400*	

Chilometri percorsi in areo in un anno

valore x_i Chilom. ↓	Freq.ass. $f_a(x_i)$	Freq. Cum. $f_{cum}(x_i)$
0	15	15
480	1	16=15+1
1250	1	17=16+1
4800	1	18=17+1
6600	1	19=18+1
9300	1	20=19+1
Tot.	*20*	

$$f_{cum}(0) = 15 = f_a(0)$$

$$f_{cum}(480) = 16=15+1 = f_{cum}(0) + f_a(480)$$

$$f_{cum}(1250) = 17=16+1 = f_{cum}(480)+f_a(1250)$$

$$f_{cum}(4800) = 18=17+1 = f_{cum}(1250)+f_a(4800)$$

$$f_{cum}(6600) = 19=18+1 = f_{cum}(4800)+ f_a(6600)$$

$$f_{cum}(9300) = 20=19+1 = f_{cum}(6600)+ f_a(9300)$$

__Calcolo delle frequenze relative cumulate ai valori effettivamente rilevati__

$F\,(x_i) = f_{cum}\,(x_i)/\,n$ ***frequenza relativa cumulata a*** x_i

Per il calcolo dei valori $F(x_i)$ ***si può ricorrere alle frequenze relative.***

Infatti, nell'ipotesi $x_1 < x_2 < \ldots < x_k$ risulta

$F\,(x_1) = f_{cum}\,(x_1)\,/\,n = f_a\,(x_1)/\,n = f_r\,(x_1)$ ***freq. rel. di*** x_1

$F\,(x_2) = f_{cum}\,(x_2)\,/\,n = (f_a\,(x_1) + f_a\,(x_2))/n = (f_a\,(x_1)/\,n) + (f_a\,(x_2)/\,n) =$

$= f_r\,(x_1) + f_r\,(x_2)$ ***somma delle freq. rel. di*** x_1 ***e freq. rel. di*** x_2

...

$F\,(x_{k-1}) = f_{cum}\,(x_{k-1})\,/\,n = f_r\,(x_1) + f_r\,(x_2) + \ldots + f_r\,(x_{k-1})$

somma delle frequenze relative di x_1***, di*** x_2 ***… e di*** x_{k-1}

$F\,(x_k) = f_{cum}\,(x_k)\,/\,n = f_r\,(x_1) + f_r\,(x_2) + \ldots + f_r\,(x_k) = 1$

somma di tutte le frequenze relative

Pertanto

Iª regola di calcolo delle frequenze relative cumulate: ***la frequenza relativa cumulata a*** x_i ***è somma (*** *quindi* ***'cumulo') delle frequenze relative di*** x_i ***e dei valori che precedono*** x_i***.***

Tabella sul kilometraggio aereo percorso in un anno

Valore x_i Chilom. ↓	freq.ass. $f_a(x_i)$	freq.rel. $f_r(x_i)$	freq.cum $f_{cum}(x_i)$	freq. rel. cum. $F(x_i)$
0	15	15/20	15	15/20
480	1	1/20	16	16/20
1250	1	1/20	17	17/20
4800	1	1/20	18	18/20
6600	1	1/20	19	19/20
9300	1	1/20	20	1
Tot.	20	1		

Da

$$F\,(x_i) = = f_r\,(x_1) + f_r\,(x_2) + \ldots + f_r\,(x_{i-1}) + f_r\,(x_i) \qquad (6)$$

segue

$$F\,(x_i) = F\,(x_{i-1}) + f_r\,(x_i) \quad \text{per } i>1 \qquad (7)$$

Pertanto:

II^a regola di calcolo delle frequenze relative cumulate. ***La frequenza relativa cumulata a un valore rilevato x_i è somma della frequenza relativa di x_i e della frequenza relativa cumulata al valore x_{i-1} immediatamente precedente.***

Tabella sul kilometraggio aereo percorso in un anno

Valore x_i Chilom. ↓	freq.ass. $f_a(x_i)$	freq.rel. $f_r(x_i)$	freq.cum $f_{cum}(x_i)$	freq. rel. Cum. $F(x_i)$
0	**15**	**15/20**	**15**	**15/20**
480	**1**	**1/20**	**16**	**16/20=(15/20)+ (1/20)**
1250	**1**	**1/20**	**17**	**17/20=(16/20)+ (1/20)**
4800	**1**	**1/20**	**18**	**18/20=(17/20)+ (1/20)**
6600	**1**	**1/20**	**19**	**19/20=(18/20)+ (1/20)**
9300	**1**	**1/20**	**20**	**1=(19/20)+ (1/20)**
Tot.	20	1		

Calcolo delle frequenze percentuali cumulate a valori rilevati

$$F_{per}\,(x_i) = (\,f_{cum}\,(x_i)/\,n\,)\times 100 = F\,(x_i)\times 100$$

Per il calcolo di $F_{per}\,(x_i)$ ***si può ricorrere alle frequenze percentuali.***

Supposto $x_1 < x_2 < \ldots < x_k$ *risulta infatti*

$$F_{per}\,(x_i) = (f_{cum}\,(x_i)/\,n)\times 100 = [(\,f_a\,(x_1) + f_a\,(x_2) + \ldots + f_a\,(x_i))/\,n]\times 100 =$$
$$= [(\,f_a\,(x_1)/\,n\,) + (f_a\,(x_2)\,/\,n) + \ldots + (f_a\,(x_i)\,/\,n)]\times 100$$
$$= (f_r\,(x_1) + f_r\,(x_2) + \ldots + f_r\,(x_i))\times 100 =$$
$$= f_r\,(x_1)\times 100 + f_r\,(x_2)\times 100 + \ldots + f_r\,(x_i)\times 100 =$$
$$= f_{per}\,(x_1) + f_{per}\,(x_2) + \ldots + f_{per}\,(x_i)$$

Pertanto

I^a^ regola di calcolo delle frequenze percentuali cumulate. ***La frequenza percentuale cumulata a*** x_1 ***è somma*** *(quindi* ***'cumulo'*** *)* ***delle frequenze percentuali di*** x_i ***e dei valori che precedono*** x_i.

$$F_{per}\,(x_i) = f_{per}\,(x_1) + f_{per}\,(x_2) + \ldots + f_{per}\,(x_{i-1}) + f_{per}\,(x_i) \qquad (8)$$

Dall'uguaglianza suindicata segue la seguente:

$$F_{perc}\,(x_i) = F_{perc}\,(x_{i-1}) + f_{perc}\,(x_i) \qquad \textit{per } i > 1 \qquad (9)$$

II^a^ regola di calcolo delle frequenze percentuali cumulate. ***La frequenza percentuale cumulata a un valore*** x_i ***è somma della frequenza percentuale di*** x_i ***e della frequenza percentuale cumulata al valore immediatamente precedente*** x_{i-1}.

Tabella su kilometraggio aereo percorso in un anno

Valore x_i Chilom. ↓	fr.ass. f_a	fr. rel. f_r	fr. perc. f_{per}	fr. cum. f_{cum}	fr.rel.cum. F	fr.per.cum F_{per}
0	15	15/20	75%	15	15/20	75%
480	1	1/20	5%	16	16/20	80%
1250	1	1/20	5%	17	17/20	85%
4800	1	1/20	5%	18	18/20	90%
6600	1	1/20	5%	19	19/20	95%
9300	1	1/20	5%	20	1	100%
Tot.	20	1	100%			

ESERCIZIO. Completare la tabella

x_i = età degli studenti	$f_a(x_i)$	$f_r(x_i)$	$f_{cum}(x_i)$	$F(x_i)$
x_1=19	880		880	
x_2=20 anni	1200		2080= 880+1200	
x_3=21 anni	600		2680= 2080+600	
x_4=22 anni	400		3080=2680+400	
x_5=23 anni	320		3400=3080+320	
	n= 3400			

3.1.1 *CALCOLO DELLE FREQUENZE CUMULATE PER OGNI x REALE E RAPPRESENTAZIONE GRAFICA DELLE FUNZIONI* f_{cum}, F e F_{perc}.

Valori x_i	$f_a(x_i)=n_i$ fr. ass. di x_i	$f_{rel}(x_i)=n_i/n$ fr. rel. di x_i	$f_{per}(x_i)= (n_i/n)\cdot 100$ fr. perc. di x_i	$f_{cum}(x_i)= n_1+\ldots+n_i$ fr. ass. cum.di x_i	$F(x_i)$ $(n_1+\ldots+n_i)/n$ Funzione di Ripartizione
x_1	n_1	n_1/n	$(n_1/n)\cdot 100$	n_1	n_1/n
x_2	n_2	n_2/n	$(n_2/n)\cdot 100$	n_1+n_2	$(n_1+n_2)/n$
…	…	…	…	…	…
x_i	n_i	n_i/n	$(n_i/n)\cdot 100$	$n_1+n_2+\ldots+n_i$	$(n_1+n_2+\ldots+n_i)/n$
…	…	…	…	…	…
x_k	n_k	n_k/n	$(n_k/n)\cdot 100$	n	$n/n=1$
	n	1	100		

La distribuzione di frequenze assolute cumulate (1)

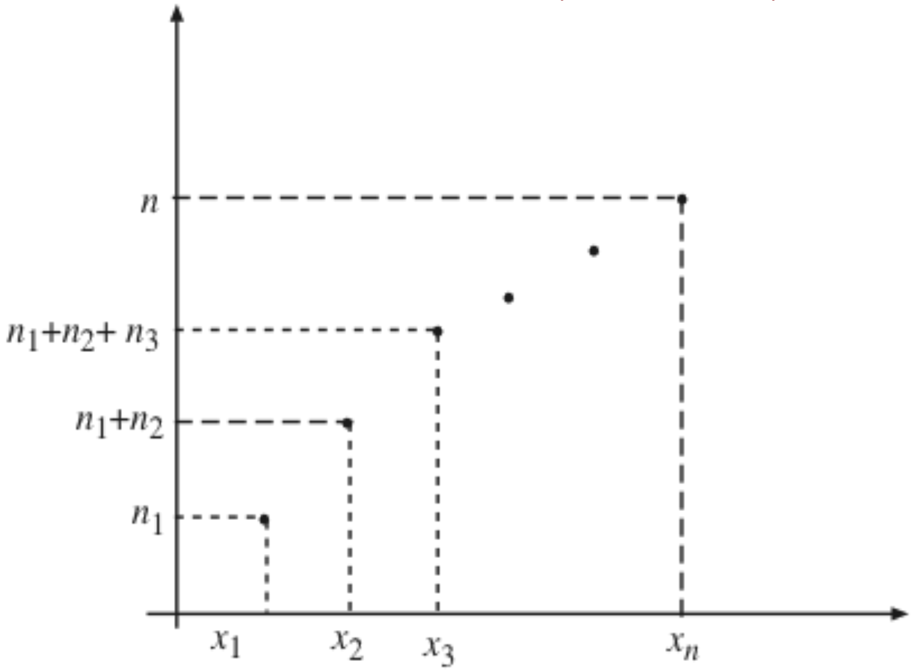
$$f_{cum}: x\in R \rightarrow f_{cum}(x)$$

è una funzione reale di una variabile reale ed è rappresentata da un grafico cartesiano.

Per determinare tutti i valori della funzione e disegnare il grafico partiamo dalla considerazione della sua restrizione all'insieme $X = \{x_1, x_2, \ldots x_k\}$ *dei valori effettivamente rilevati:*

$$(f_{cum})_X : x_i \in X = \{x_1, x_2, \ldots x_k\} \rightarrow f_{cum}(x_i) = (n_1+\ldots+n_i)$$

Tale restrizione è una funzione strettamente crescente che ha per grafico ***l'insieme dei punti*** • del piano di coordinate $(x_i, f_{cum}(x_i)) = (x_i, n_1+\ldots+n_i)$ *(*)*



(*) nel grafico l'ultimo punto segnato sull'asse delle ascisse è x_k *e non* x_n

Calcoliamo i valori di f_{cum} su tutto l'insieme R dei numeri reali.

Conoscendo le frequenze cumulate ai valori $x_1, x_2, \ldots, x_k$, $(x_1 < x_2 < \ldots < x_k)$, possiamo calcolare **la** frequenza cumulata a un qualsiasi $x \in \mathbf{R}$, cioè rispondere alla domanda:

d) quale è il numero dei casi in cui il carattere assume valore ≤ x?

Sia $x < x_1$:

$x \quad x_1 \quad x_2 \quad x_3 \qquad\qquad\qquad x_k$

Poiché non stati rilevati valori minori di x_1, ***0 è il numero dei casi con valori rilevati minori di x o uguali a x,*** *allora poniamo* $f_{cum}(x) = 0$.

Sia $x \in [x_1, x_2[$:

$x_1 \quad x \quad x_2 \qquad x_3$

Poiché il carattere non assume valori compresi strettamente tra x_2 **e** x_3, *la risposta alla domanda d) è* n_1, *numero dei casi in cui il carattere assume il valore* x_1, ***unico valore rilevato ≤ x:***

$$f_{cum}(x) = n_1 = f_{cum}(x_1)$$

Sia $x \in [x_2, x_3[$:

$x_1 \quad x_2 \quad x \quad x_3$

Poiché il carattere non assume valori compresi strettamente tra x_2 **e** x_3, ***i casi con valore del carattere ≤x coincidono con i casi con valore ≤*** x_2, *pertanto la risposta è* n_1+n_2, *dove* n_1 *è il numero dei casi in cui il carattere assume valore* x_1 *e* n_2 *è il numero dei casi in cui il carattere assume valore* x_2; *poniamo allora*

$$f_{cum}(x) = n_1+n_2 = f_{cum}(x_2).$$

In generale*: non ci sono valori del carattere strettamente compresi tra i valori successivi rilevati* x_i **e** x_{i+1}; *pertanto se* $x_i < x < x_{i+1}$ *il numero dei casi con valore del carattere* **≤ x** *coincide con il numero di casi con valore ≤* x_i *ed pari quindi a* $n_1+\ldots+n_i$.(*) *Se invece* $x \geq x_k$, *essendo* x_k *il più grande valore rilevato, tutti gli* **n** *casi considerati presentano valore del carattere* **≤ x.**

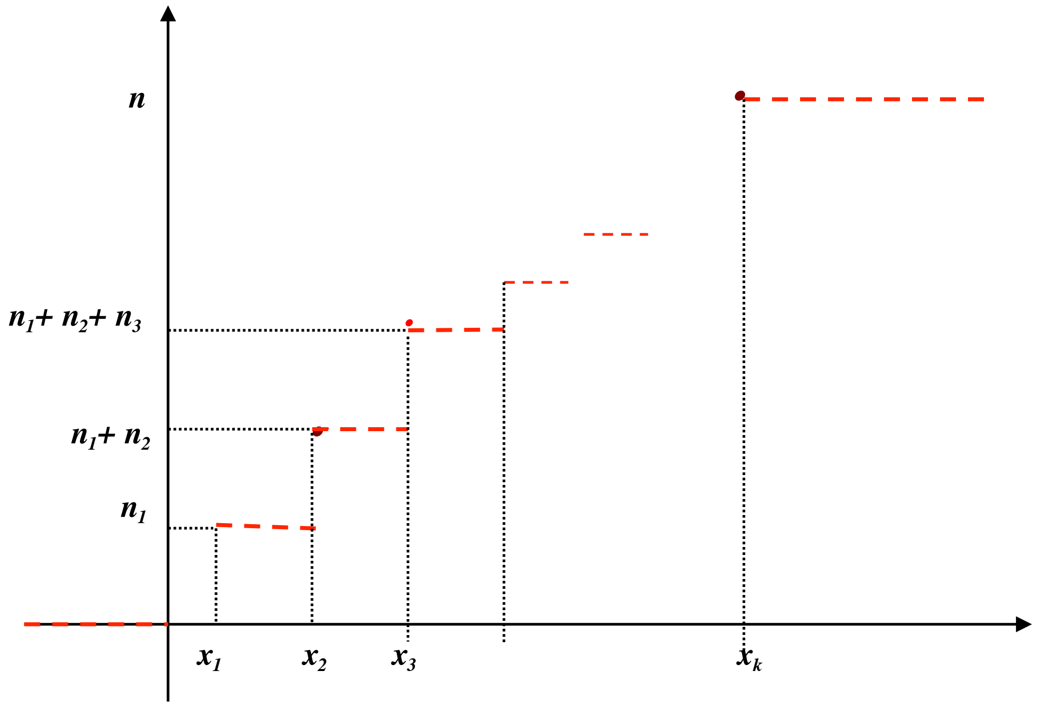
In definitiva:

$f_{cum}(x) = 0$	**per** $x < x_1$,	
…		
$f_{cum}(x) = f_{cum}(x_i)$	**per** $x \in [x_i, x_{i+1}[$	$i=1, 2, \ldots, k-1$
…		
$f_{cum}(x) = f_{cum}(x_k) = n$	**per** $x \geq x_k$.	

f_{cum} è allora ***la*** **la funzione costante a tratti:**

$$f_{cum}: x \in R \to \begin{cases} 0 & se & x < x_1 \\ n_1 & se & x \in [x_1, x_2[\\ \vdots & \vdots & \vdots \\ n_1 + n_2 + \ldots + n_i & se & x \in [x_i, x_{i+1}[\\ \vdots & \vdots & \vdots \\ n_1 + n_2 + \ldots + n_k = n & se & x \geq x_k \end{cases}$$

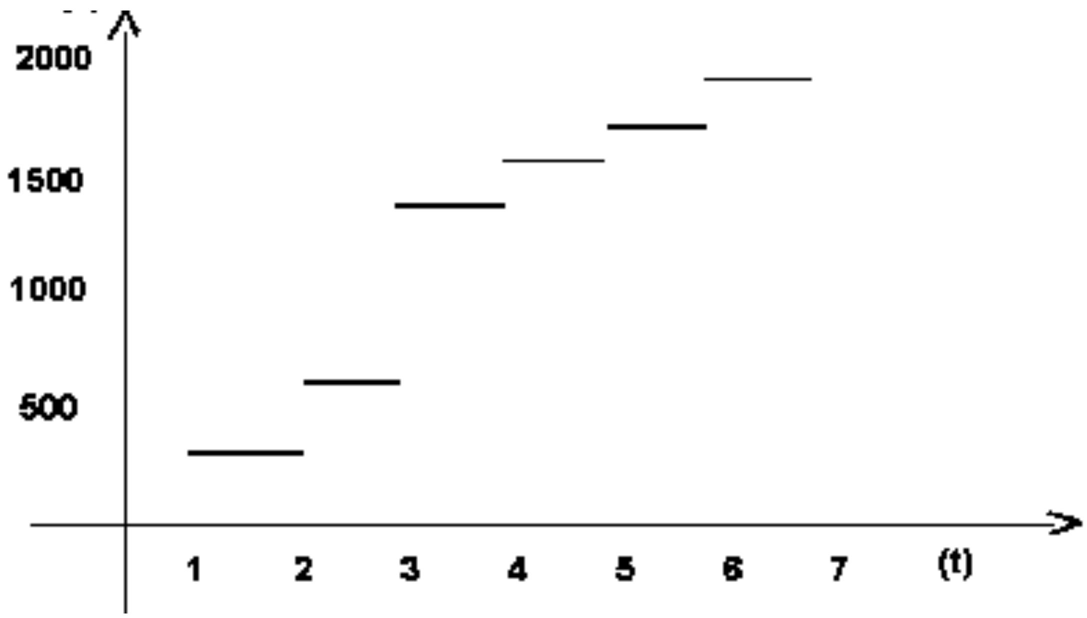
Il grafico cartesiano é un grafico a scalini rappresentato in figura dai segmenti rossi chiusi a sinistra



La funzione f_{cum} , risultando costante in ognuno degli intervalli]-∞, x_1[, [x_1, x_2[, ….,[x_{k-1}, x_k[, [x_k, +∞ [, è solo **crescente.**

Esempio: giorni occorrenti a un plico di lettere per raggiungere l'ufficio destinatario.

Giorni (x)	Frequenze	Frequenze cumulate
1	157	157
2	419	576
3	890	1457
4	215	1672
5	120	1792
6	150	1942
7	95	2037
	2037	



Calcoliamo i valori della funzione di ripartizione F **su tutto l'insieme R dei numeri reali.**

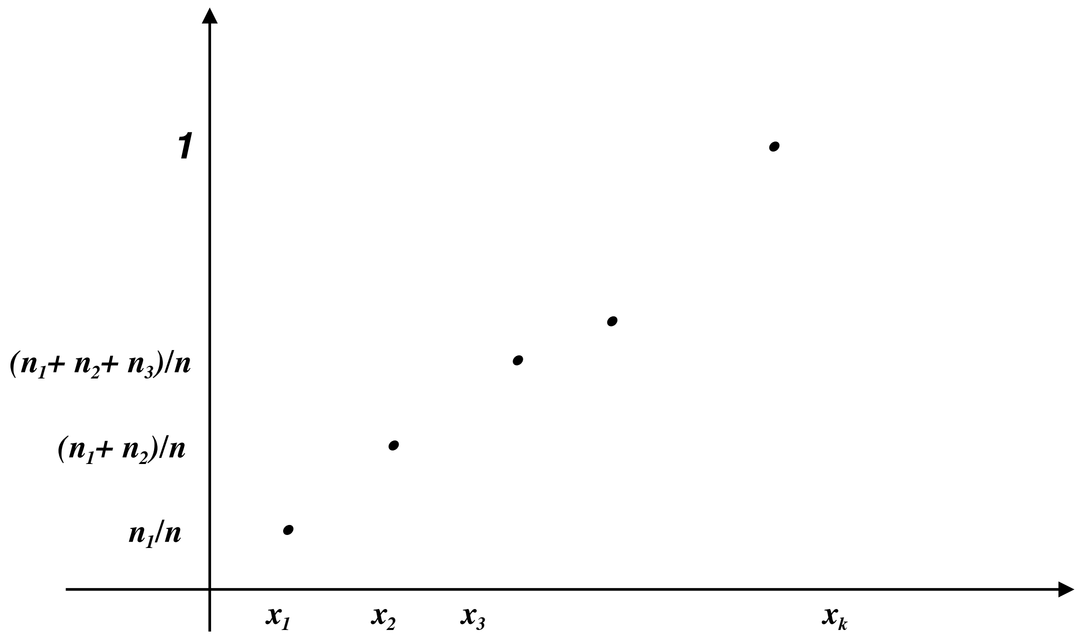
La restrizione F all'insieme $X = \{ x_1, x_2, \ldots, x_k \}$ *dei valori rilevati è*

F_X: $x_i \in \{x_1, x_2, \ldots, x_k\} \rightarrow F(x_i) = (n_1 + \ldots + n_i)/n$ *Funzione di ripartizione su X*

e assume i valori:

$F(x_1) = n_1/n$, $F(x_2) = (n_1+n_2)/n$, …, $F(x_i) = (n_1+\ldots+n_i)/n$, …, $F(x_k) = (n_1+\ldots+n_k)/n = 1$

F_X *è una funzione* **strettamente crescente** *il cui grafico è indicato in figura*



Costruiamo la funzione di ripartizione da F_X . Ragionando come nel caso della distribuzione di frequenze assolute cumulate poniamo:

$F(x) = 0$ se $x < x_1$

$F(x) = F(x_i) = (n_1+\ldots+n_i)/n$ se $x_i \leq x < x_{i+1}$

$F(x) = 1$ se $x \geq x_k$ ***massimo dei valori rilevati***

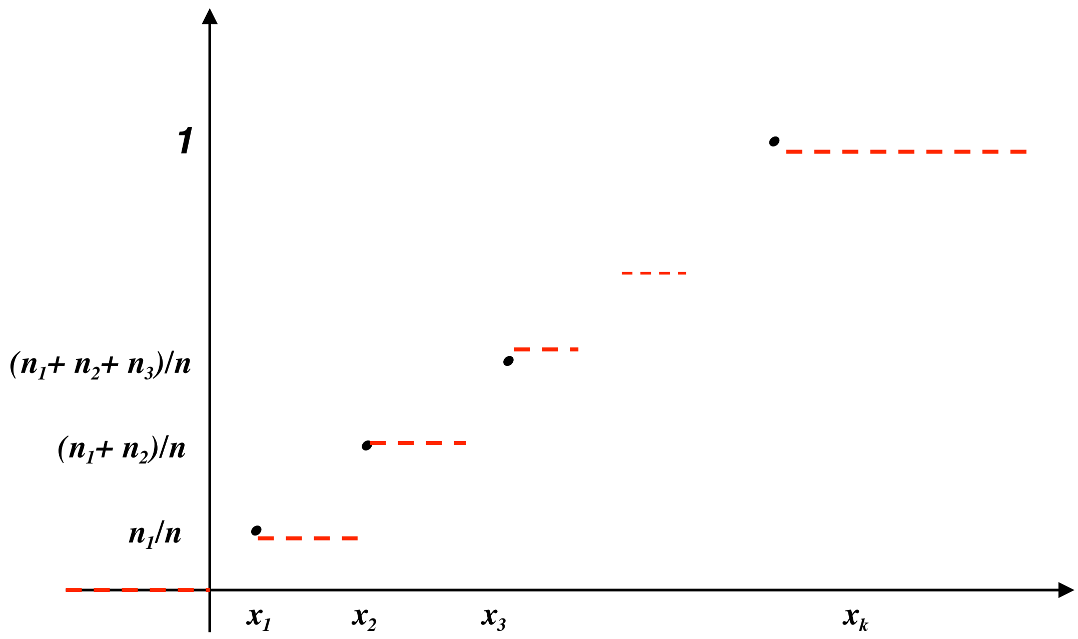
Allora **la funzione di ripartizione,**

$$F: x \in \,]-\infty, +\infty[\, \rightarrow F(x) = (\text{num. casi con valore} \leq x)/n =)\ \overline{f}_{cum}(x)\ /n$$

è la funzione costante a tratti

$$F : x \in \left]-\infty,+\infty\right[\rightarrow F(x) = \begin{cases} 0 & se\ x < x_1 \\ n_1/n & se\ x \in [x_1, x_2[\\ \vdots & \vdots \\ (n_1 + n_2 + \ldots + n_i)/n & se\ x \in [x_i, x_{i+1}[\\ \vdots & \vdots \\ (n_1 + n_2 + \ldots + n_k)/n = 1 & se\ x \geq x_k \end{cases}$$

con valore minimo 0 e valore massimo 1.
***Il suo grafico si presenta sotto forma di* grafico a scala.**



La distribuzione di frequenze percentuali cumulate, F_{perc}, viene determinata allo stesso modo e presenta anche essa un grafico a scala: il suo valore massimo è100

3.2 *CALCOLO DELLE FREQUENZE CUMULATE PER IL CARATTERE QUANTITATIVO CONTINUO CON VALORI RAGGRUPPATI IN CLASSI E* **OGIVA O POLIGONO DELLE FREQUENZE CUMULATE**

Ricordiamo la definizione di frequenza assoluta cumulata da noi considerata (una diversa definizione è data in appendice)

$f_{cum}(x)$ = numero delle unità con determinazione carattere minore o uguale a x

Per utilizzare tale definizione nel modo migliore conviene supporre che

i valori siano raggruppati in classi contigue chiuse a destra

Se i valori del carattere sono compresi tra i valori x_0 e x_k, le classi sono del tipo

$$x_0 —| x_1,\ \ x_1 —| x_2,\ \ x_2 —| x_3,\ \ \ldots,\ \ \ x_{k-1} —| x_k,$$

dove si suppone che x_0 non sia un valore preso e che x_k sia il massimo dei valori presi; altrimenti la prima classe può la classe aperta " $\leq x_0$".

Le uniche frequenze cumulate che possiamo allora calcolare sono quelle corrispondenti agli estremi superiori delle classi indicate secondo l'ordine crescente degli estremi. Prendiamo ad esempio la tabella

Classi di valori $]x_{i-1}, x_i]$	f_a	$f_{cum}(x_i)$
x_0=10—| 20 = x_1	7	7= $f_{cum}(20)$
20—| 30= x_2	4	11=7+4 = $f_{cum}(30)$
30—| 40= x_3	5	16=7+4+5 = $f_{cum}(40)$
40— | 50= x_4	2	18=7+4+5+2 = $f_{cum}(50)$
50— | 60= x_5	2	20=7+4+5+2+2 = $f_{cum}(60)$
	20	

La frequenza assoluta **7** corrispondente alla prima classe **10— | 20** indica anche il numero delle unità con determinazione del carattere minore o uguale a **20 e** quindi è **7 =** $f_{cum}(20)$; analogamente la somma delle frequenze corrispondenti alle prime due classi**, 11=7+4**, indica il numero delle delle unità con determinazione del carattere minore o uguale a **30** ed è quindi **11=** $f_{cum}(30)$; e così procedendo la somma delle frequenze delle prime **'*i*'** classi è la frequenza cumulata $f_{cum}(x_i)$ all'estremo superiore l_i della ***i***-esima classe**.**

Sappiamo allora solo costruire la seguente funzione che è ***la restrizione della distribuzione delle frequenze assolute cumulate all'insieme degli estremi superiori delle classi di valori***

$$
f^{*}_{cum}: \quad
\begin{aligned}
x_1 &= 20 \rightarrow 7 = f_{cum}(20) \\
x_2 &= 30 \rightarrow 11 = f_{cum}(30) \\
x_3 &= 40 \rightarrow 16 = f_{cum}(40) \\
x_4 &= 50 \rightarrow 18 = f_{cum}(50) \\
x_5 &= 60 \rightarrow 20 = f_{cum}(60)
\end{aligned}
$$

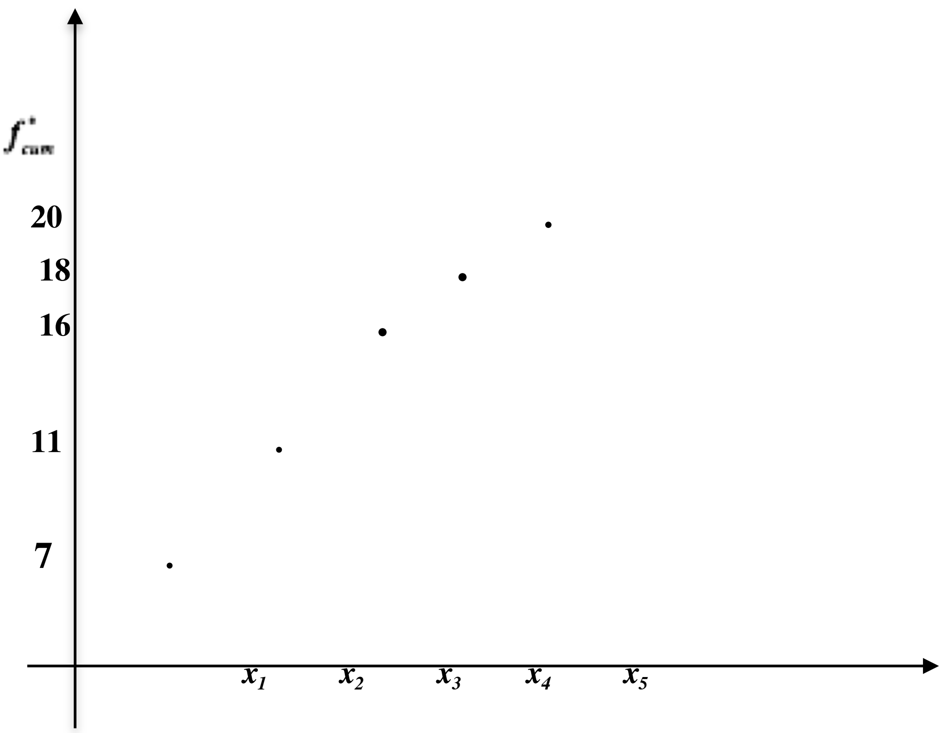


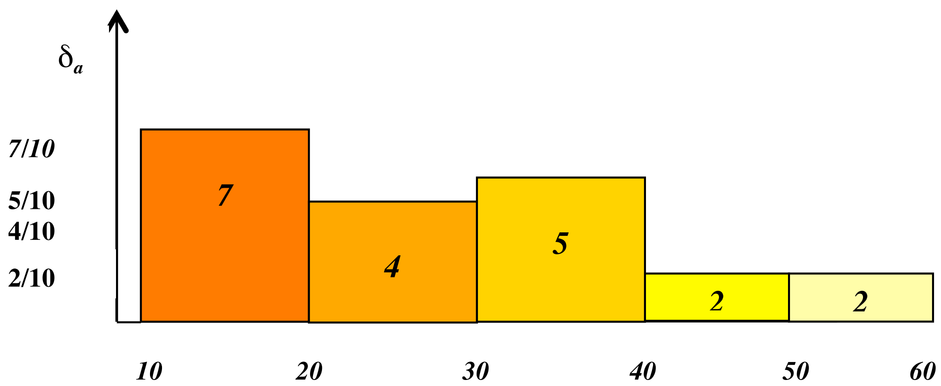
grafico cartesiano di f^{*}_{cum}

NOTA I valori della funzione costruita

$$f^*_{cum} : \begin{array}{l} x_1 = 20 \rightarrow 7 \\ x_2 = 30 \rightarrow 11 \\ x_3 = 40 \rightarrow 16 \\ x_4 = 50 \rightarrow 18 \\ x_5 = 60 \rightarrow 20 \end{array}$$

si possono leggere anche attraverso ***l'istogramma delle frequenze assolute*** ottenuto riportando gli intervalli consecutivi che rappresentano le classi con i loro limiti reali e costruendo su ogni intervallo un rettangolo di altezza pari alla densità di frequenza assoluta corrispondente alla classe rappresentata dall'intervallo.

Classi di valori $]x_{i-1}, x_i]$	f_a	**δ = dens. di freq. ass.**	$f_{cum}(x_i)$
$x_0=10$ —| $20 = x_1$	**7**	**7/10=0,7**	**7=** $f_{cum}(20)$
20—| $30= x_2$	**4**	**4/10=0,4**	**11=7+4 =** $f_{cum}(30)$
30—| $40= x_3$	**5**	**5/10=0,5**	**16=7+4+5 =** $f_{cum}(40)$
40— | $50= x_4$	**2**	**2/10=0,2**	**18=7+4+5+2 =** $f_{cum}(50)$
50— | $60= x_5$	**2**	**2/10=0,2**	**20=7+4+5+2+2 =** $f_{cum}(60)$



l'area del primo rettangolo è $f_{cum}(20)$;
l'area della figura costituita dai primi due rettangoli è $f_{cum}(30)$
l'area della figura costituita dai primi tre rettangoli è $f_{cum}(40)$
…………..
l'area dell'istogramma è $f_{cum}(60) = 20$

Calcolo dei valori di f_{cum} su tutto R.

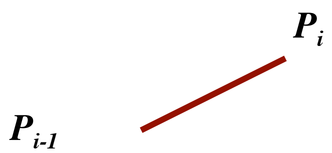
Sia $x < x_0$; con il ragionamento seguito nel caso del carattere discreto, possiamo affermare che $f_{cum}(x) = 0$.

Sia $x \geq x_k$; seguendo il ragionamento utilizzato nel caso del carattere discreto, possiamo anche ora affermare che $f_{cum}(x) = n$ (numero di tutte le unità statistiche).

Sia $x \in [x_i, x_{i+1}[$: non possiamo ora affermare che la frequenza cumulata a x sia uguale a quella cumulata a x_i, perché, contrariamente con quello che accade nel caso discreto, ci sono valori tra x_i e x, che possono essere presi dal carattere e allora $f_{cum}(x_i) < f_{cum}(x)$: in questo caso, poiché non sappiamo quali valori sono presi e con quale frequenza, siamo nell'impossibilità di valutare $f_{cum}(x)$.

Un caso particolare: *la distribuzione delle frequenze è <u>uniforme</u> all'interno di ogni classe di valori; allora il grafico di f_{cum} su ogni intervallino $]x_{i-1}, x_i]$ è il segmento congiungente i punti*

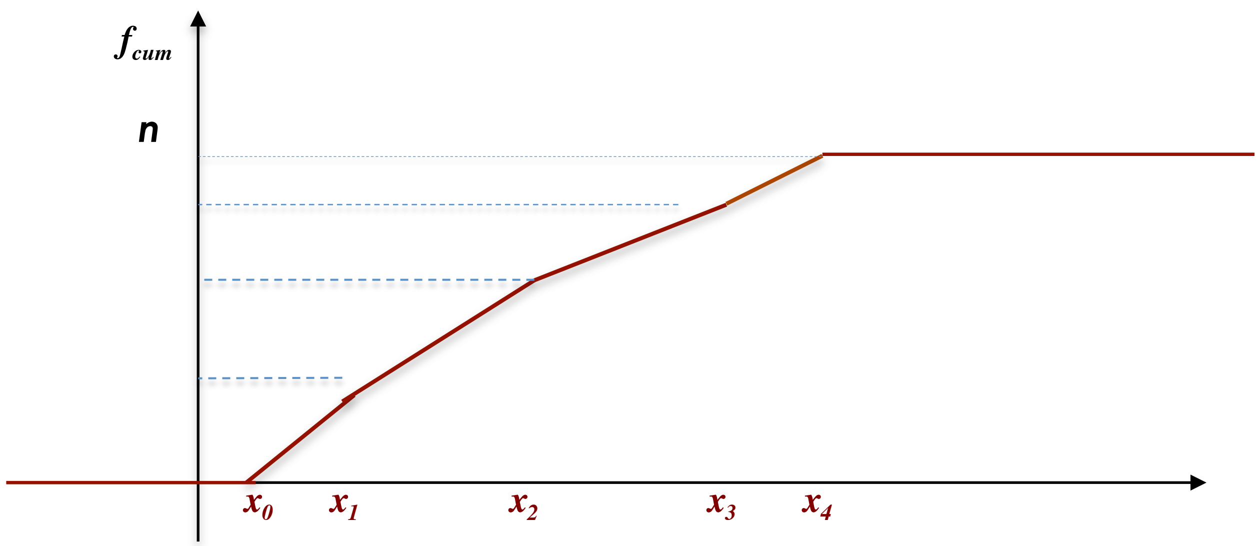
$$P_{i-1}(x_{i-1}, f_{cum}(x_{i-1})) \qquad e \qquad P_i(x_i, f_{cum}(x_i))$$



di equazione:

$$y = f_{cum}(x_{i-1}) + \frac{f_{cum}(x_i) - f_{cum}(x_{i-1})}{x_i - x_{i-1}}(x - x_{i-1})$$

Allora
il grafico è una spezzata detta **"ogiva o poligono delle frequenze assolute cumulate".**



Ogiva o poligono delle frequenze assolute cumulate

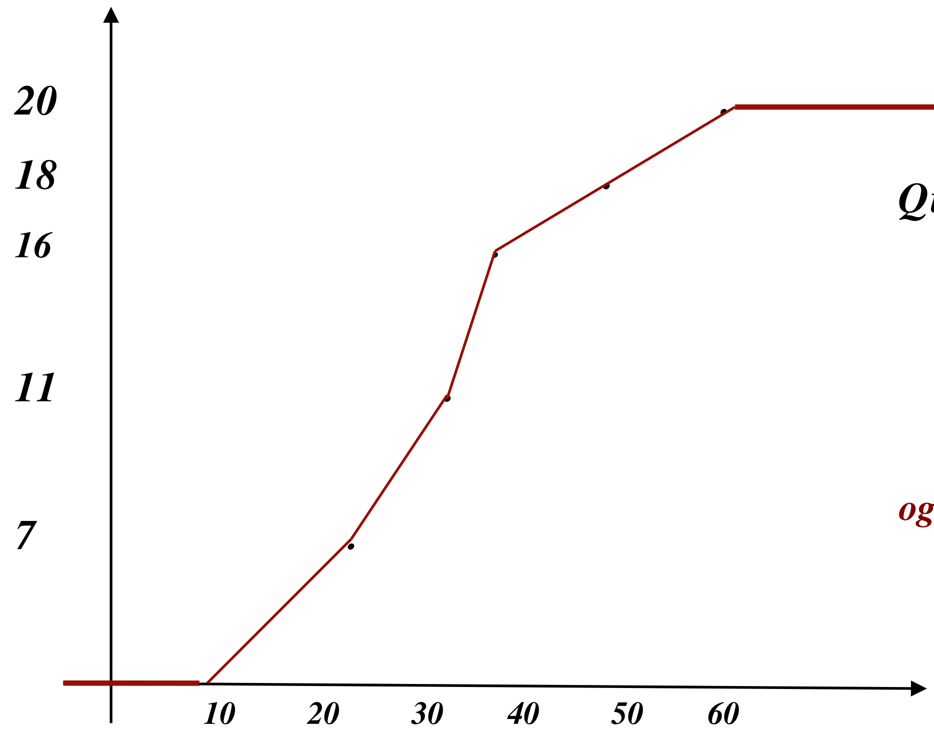
Esempio

Classi di valori $]x_{i-1}, x_i]$	f_a	$f_{cum}(x_i)$
$x_0=10$ —| $20 = x_1$	**7**	$7 = f_{cum}(20)$
20 —| $30 = x_2$	**4**	$11 = 7+4 = f_{cum}(30)$
30 —| $40 = x_3$	**5**	$16 = 7+4+5 = f_{cum}(40)$
40 —| $50 = x_4$	**2**	$18 = 7+4+5+2 = f_{cum}(50)$
50 —| $60 = x_5$	**2**	$20 = 7+4+5+2+2 = f_{cum}(60)$
	20	

$f_{cum}: x \in \mathbf{R} \rightarrow f_{cum}(x) =$ ***numero dei casi con determinazione del car.*** $\leq x$

distribuzione di frequenze cumulate

Costruiamo prima il grafico *sul range dei valori* dei valori **]10, 60]**. Ponendo $f_{cum}(10) = 0$ (essendo la prima classe aperta a 10 non ci sono unità con determinazione del carattere ≤ 10) e supponendo che la restrizione della la funzione ad ognuno degli intervalli *[10, 20], [20, 30], [30, 40], [40, 50], [50, 60],* sia affine otteniamo il seguente grafico . Ponendo quindi $f_{cum}(x) = 0$ per $x<10$ e $f_{cum}(x) = 20$ per $x > 60$, ***tteniamo il grafico seguente***



Quanto vale $f_{cum}(35)$***?***

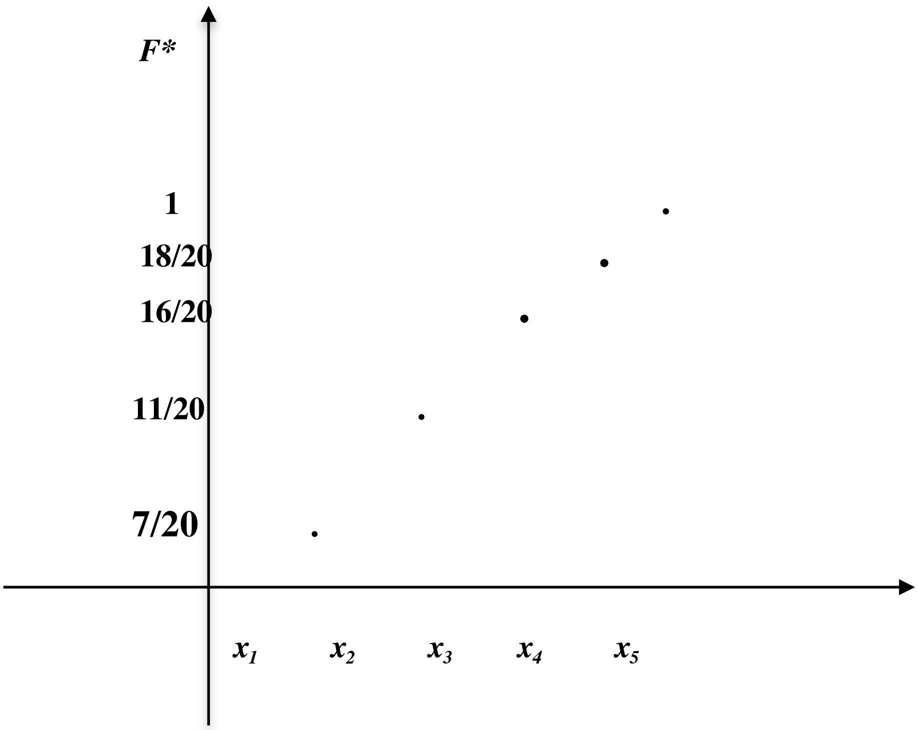
ogiva delle frequenze cumulate

Sui valori di *F*.

Un analogo discorso vale per le **distribuzioni** di **frequenze relative cumulate** o di **frequenze percentuali cumulate**: la frequenza relativa (*risp.* percentuale) corrispondente alla prima classe è la frequenza relativa (risp. percentuale) cumulata $F(x_1)$ (*risp.* $F(x_1) \times 100$) all'estremo superiore x_1 della prima classe**;** la somma delle frequenze relative (risp. percentuali) delle prime 'i' classi è la frequenza cumulata $F(x_i)$ (*risp.* $F(x_i) \times 100$) all'estremo superiore x_i della i-esima classe**.**

Classi di valori $]x_{i-1}, x_i]$	f_a	f_r	f_{per}	F	F_{per}
10— | 20	7	7/20=0,35	35%	7/20 = $F(20)$	35%
20— | 30	4	4/20=0,2	20%	11/20= $F(30)$	55%
30— | 40	5	5/20=0,25	25%	16/20= $F(40)$	80%
40— | 50	2	2/20=0,1	10%	18/20= $F(50)$	90%
50— | 60	2	2/20=0,1	10%	20/20= $F(60)$=1	100%
	20	1	100%		

F* restrizione della funzione di ripartizione all'insieme costituito dagli estremi superiori delle classi

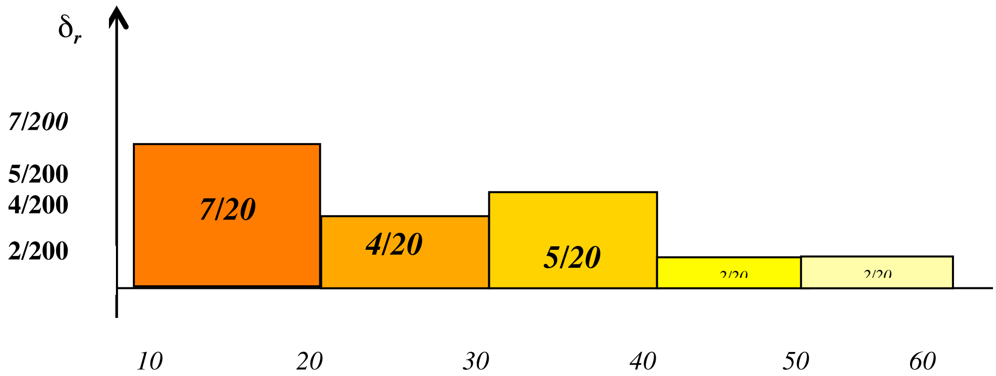


NOTA I valori della funzione

$$F^*: \begin{array}{l} x_1 = 20 \rightarrow 7/20 = \\ x_2 = 30 \rightarrow 11/20 \\ x_3 = 40 \rightarrow 16/20 \\ x_4 = 50 \rightarrow 18/20 \\ x_5 = 60 \rightarrow 20/20 = \end{array}$$

rappresentata prima graficamente si possono leggere anche attraverso ***l'istogramma delle frequenze relative*** ottenuto riportando gli intervalli consecutivi che rappresentano le classi con i loro limiti reali e costruendo su ogni intervallo un rettangolo di altezza pari alla densità di frequenza relativa corrispondente alla classe rappresentata dall'intervallo.

Classi di valori $]x_{i-1}, x_i]$	$\mathbf{f_a}$	$\mathbf{f_r}$	δ_a **= dens. di frequenza ass.**	δ_r **= dens. di frequenza rel.**	F
10—|20	**7**	**7/20=0,35**	**7/10=0,7**	**7/200=0,035**	**7/20 =** $F(20)$
20—|30	**4**	**4/20=0,2**	**4/10=0,4**	**4/200=0,02**	**11/20=** $F(30)$
30—|40	**5**	**5/20=0,25**	**5/10=0,5**	**5/200=0,025**	**16/20=** $F(40)$
40—|50	**2**	**2/20=0,1**	**2/10=0,2**	**2/200=0,01**	**18/20=** $F(50)$
50—|60	**2**	**2/20=0,1**	**2/10=0,2**	**2/200=0,01**	**20/20=** $F(60)$**=1**
	20	**1**			



l'area del primo rettangolo è $F(20)$;
l'area della figura costituita dai primi due rettangoli è $F(30)$
l'area della figura costituita dai primi tre rettangoli è $F(40)$

……………………

l'area dell'istogramma è 1

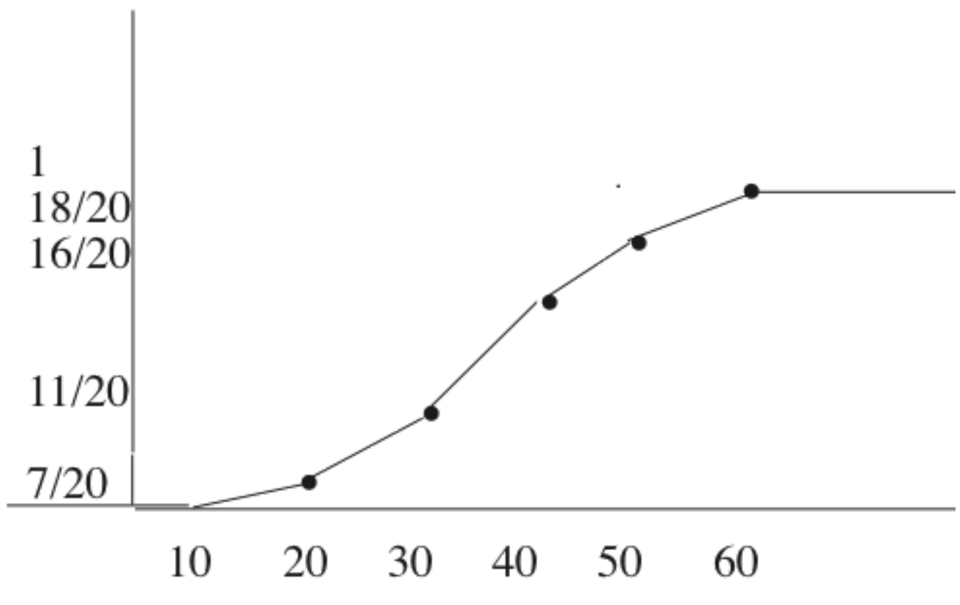
Calcoliamo i valori di F su tutto R in caso di distribuzione uniforme.

L' assunzione di uniformità della distribuzione di frequenze su ogni intervallino $]x_{i-1}, x_i]$ *permette di determinare i valori della funzione di ripartizione F a partire dai valori assunti negli estremi delle classi. Si pone*

$F(x)= 0$ *per* $x < x_0$ *ed* $F(x)= 1$ *(frequenza totale) per* $x > x_k$

mentre sull'intervallo $]10, 60]$ *, i valori sono assegnati in modo che il grafico sia costituito dai segmenti di estremi* $P_{i-1}(x_{i-1}, F(x_{i-1}))$ *e* $P_i(x_i, F(x_i))$ *ognuno dei quali ha equazione…………….*

Classi di valori $]x_{i-1}, x_i]$	f_a	f_r	F
10— |20	7	7/20=0,35	7/20 = $F(20)$
20— |30	4	4/20=0,2	11/20= $F(30)$
30— |40	5	5/20=0,25	16/20= $F(40)$
40— |50	2	2/20=0,1	18/20= $F(50)$
50— |60	2	2/20=0,1	20/20= $F(50)$=1
	20	1	



ogiva delle frequenze relative cumulate

In modo analogo si procede per la definizione e la rappresentazione grafica della distribuzione di frequenza percentuale su tutto R

IL CASO IN CUI LE CLASSI DI VALORI SONO INDICATE ATTRAVERSO LIMITI FORMALI E NON REALI

Per determinare le funzioni di distribuzioni e di ripartizione come funzioni di una variabile reale occorre riferirsi ai limiti superiori reali delle classi

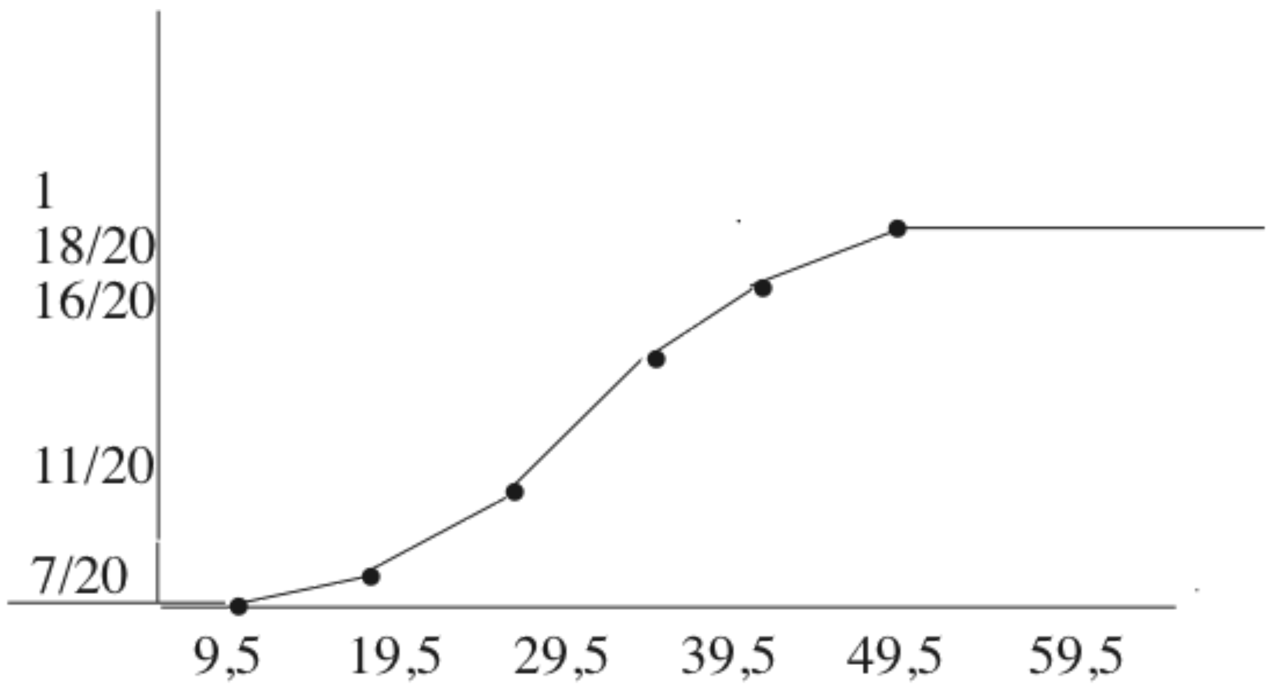
Classi di valori	f_a	f_{cum}	F
10— 19	7	7	7/20=0, 35
20— 29	4	11	11/20=0,55
30— 39	5	16	16/20=0,8
40— 49	2	18	18/20=0,9
50— 59	2	20	1
	20		

Passiamo dalla tabella data a quella in cui le classi sono indicate attraverso i confini reali

Classi di valori	f_a	f_{cum}	F
9,5— | 19,5	7	7	7/20=0,35
19,5— | 29,5	4	11	11/20=0,55
29,5—| 39,5	5	16	16/20=0,8
39,5— | 49,5	2	18	18/20=0,9
49,5— | 69,5	2	20	1
	20		

Classi di valori	f_a	f_{cum}	F
9,5—|19,5	7	7	7/20=035
19,5—|29,5	4	11	11/20=0,55
29,5—|39,5	5	16	16/20=0,8
39,5—|49,5	2	18	18/20=0,9
49,5—|69,5	2	20	1
	20		

$$F: x \in R \rightarrow f_{cum}(x)/20$$



Ogiva di *F*

3.3 DISTRIBUZIONI DI FREQUENZE CUMULATE PER UN CARATTERE QUALITATIVO ORDINABILE

Se un carattere è **qualitativo ordinato** ***le sue modalità possono essere disposte in livelli di una gerarchia***; ha senso allora chiedersi quante sono le unità statistiche in cui la determinazione del carattere è in un livello che non supera un livello fissato.

Ordinate le modalità secondo i livelli della gerarchia

$$m_1, m_2, \dots, m_k \qquad (10)$$

si definisce **frequenza cumulata** ***alla modalità*** m_i ***il numero delle unità statistiche per le quali la determinazione del carattere non supera*** m_i ***, cioè la modalità è*** m_i ***o precede*** m_i ***nell' allineamento (10).***

$f_{cum}(m_i)$= ***n. unità con determinazione del carattere che non supera*** m_i

Risulta:

$f_{cum}(m_1) = f_a(m_1)$

$f_{cum}(m_2) = f_a(m_1) + f_a(m_2) = f_{cum}(m_1) + f_a(m_2)$

$f_{cum}(x_3) = f_a(m_1) + f_a(m_2) + f_a(m_3) = f_{cum}(m_2) + f_a(m_3)$

……

$f_{cum}(m_k) = f_a(m_1) + f_a(m_2) + \dots + f_a(m_k) = f_{cum}(m_{k-1}) + f_a(m_k) = n$

Tot. unità stat.

La funzione

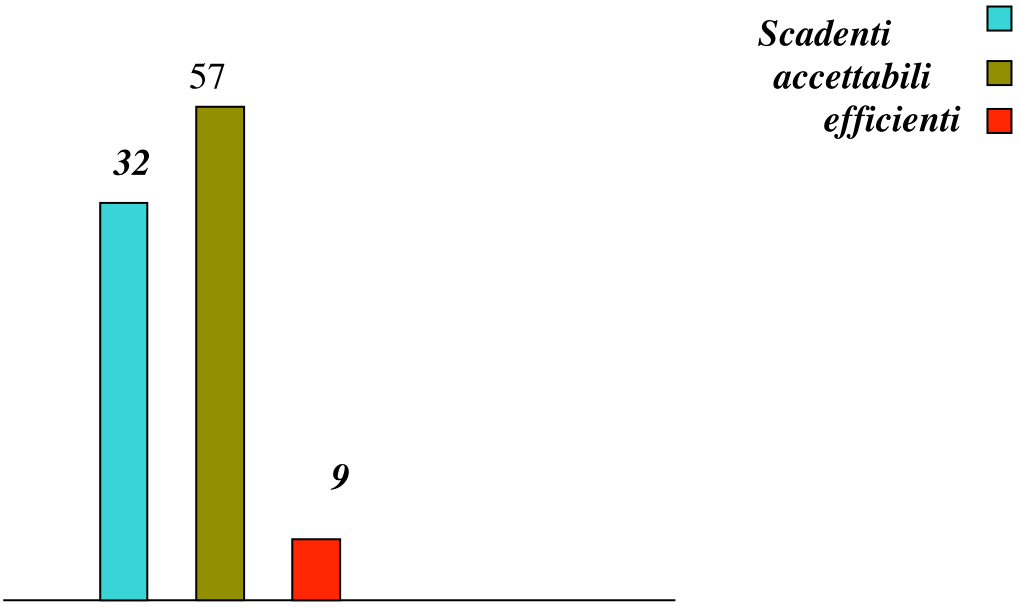
$f_{cum}: m_i \in M = \{ m_1, m_2, \dots, m_k \} \rightarrow f_{cum}(m_i)$

è detta **distribuzione semplice di frequenza cumulata**

Esempio da C. Rossi " la matematica dell'incertezza" Zanichelli

Nella seguente tabella si riporta il giudizio di 98 persone sulla condizioni dei mezzi di trasporto: i giudizi sono ordinati in una gerarchia di livelli per consentire il calcolo delle frequenze cumulate

Modalità	$f_a(m)$	$f_{cum}(m)$
Scadenti	***32***	***32***
Accettabili	***57***	***89=32+57***
Efficienti	***9***	***98=32+57+9= 89+9***
Tot.	*98*	



Possiamo leggere le frequenze cumulate attraverso le aree dei rettangoli dell'ortogramma?

Nota: *Assegnando un codice a ogni modalità, che indichi il livello occupato nella gerarchia dei giudizi*

Scadenti=1 ***Accettabili=2*** ***Efficienti =3***

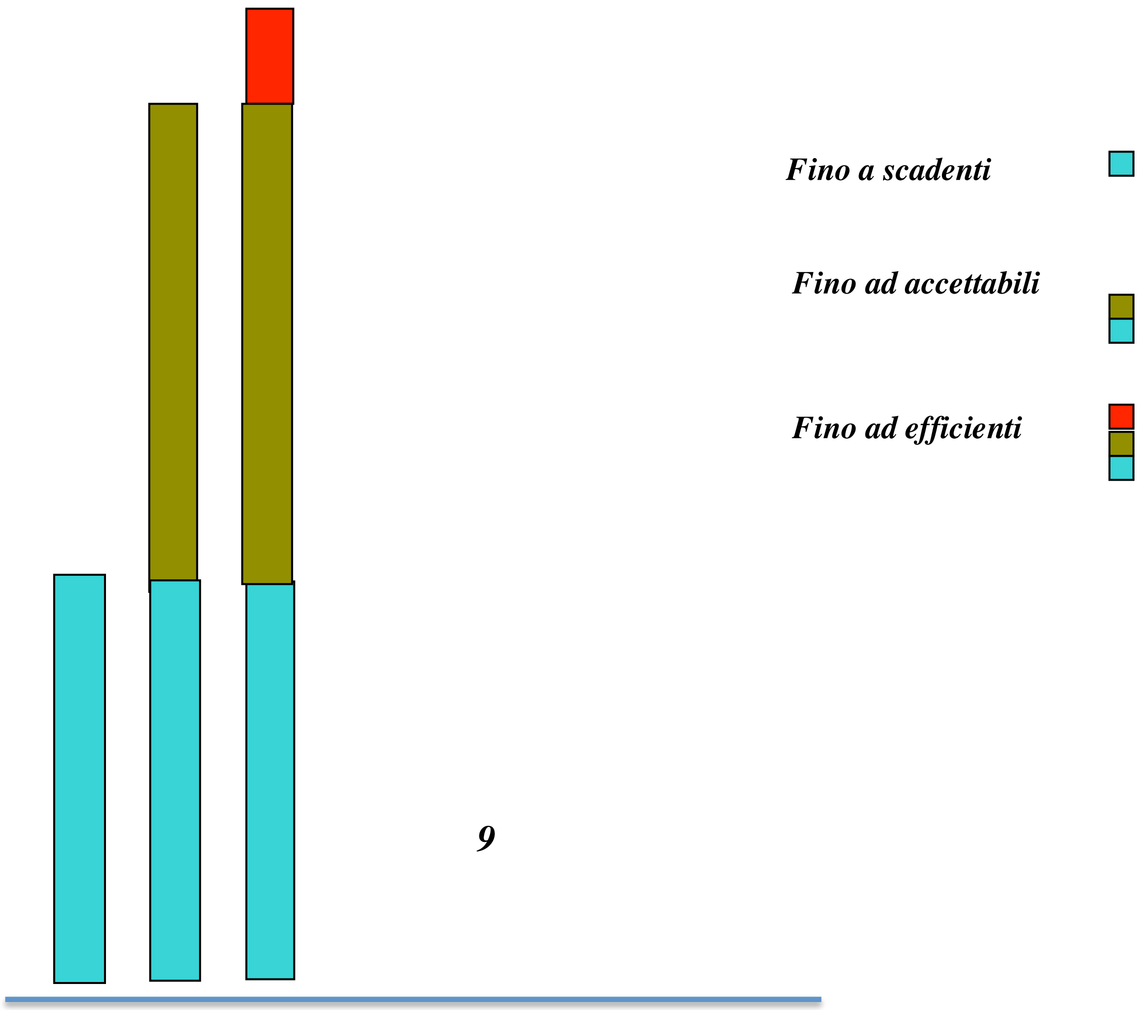
la tabella può essere cosi riscritta

Modalità	$f_a(x)$	$f_{cum}(x)$
1	***32***	***32***
2	***57***	***89***
3	***9***	***98***
Tot.	*98*	

La distribuzione di frequenze cumulate è allora la seguente

$$f_{cum}: \begin{array}{ccc} scadenti & \rightarrow & 32 \\ accettabili & \rightarrow & 89 \\ efficienti & \rightarrow & 98 \end{array}$$

e per la sua rappresentazione grafica utilizziamo ancora l'ortogramma.



In modo analogo si definiscono le frequenze relative cumulate e percentuali cumulate

Esempio da C. Rossi " la matematica dell'incertezza" Zanichelli

Tab. 4.2 Distribuzione statistica relativa alla variabile ordinata " condizione dei mezzi di trasporto"

Modalità	$n = f_a(m)$	f_r	f_{cum}	F	F_{per}
Scadenti	***32***	***0,3265***	***32***	***0,3265***	
Accettabili	***57***	***0,5816***	***89=32+57***	***0,9081=0,3265+0,5816***	
Efficienti	***9***	***0,0918***	***98=32+57+9***	***1=0,3265+0,5816+0,0918***	
Tot.	*98*	*1*			

Scadenti=1
Accettabili=2
Efficienti =3

Modalità	$n = f_a$	f_r	f_{cum}	F
1	***32***	***0,3265***	***32***	***0,3265***
2	***57***	***0,5816***	***89***	***0,9081***
3	***9***	***0,0918***	***98***	***1***
Tot.	*98*	*1*		

3.4 DISTRIBUZIONI DI QUANTITÀ CUMULATE

Partendo dalle distribuzioni di quantità o di intensità relative ad un carattere quantitativo o qualitativo ordinabile si possono definire le distribuzioni di quantità cumulate

Al concorso XXXX sono accettate le domande arrivate tra il lunedì e il venerdì di una data settimana e il giorno è indicato dal timbro di accettazione

La seguente tabella indica il numero di lettere giunte giorno per giorno

Giorno di presentazione	q_a= n. domande arrivate nel giorno indicato
lunedì	7
martedì	8
Mercoledì	10
Giovedì	12
venerdi	23
totale	60

Collettivo= insieme delle domande presentate per il concorso X

Carattere ordinato = giorno di arrivo

Si vuole conoscere per ogni giorno della settimana il numero di domande arrivate <u>entro</u> quel giorno

Le domande entro lunedì sono 7
Le domande arrivate entro martedì sono 7+ 8=15
Le domande arrivate entro martedì sono 7+ 8 + 10=25
…………………
…………………

La tabella va completata nel seguente modo

Giorno di presentazione	q_a= n. domande arrivate nel giorno indicato	$Q=q_{cum}$= n. domande arrivate entro il giorno indicato
lunedì	7	7
martedì	8	15
Mercoledì	10	25
Giovedì	12	37
venerdi	23	60
totale	60	

I numeri indicati nell'ultima colonna vengono dette
Quantità o intensità cumulate

7= n. domande cumulate il primo giorno =q_{cum}(lunedì)

15= 7+8 domande cumulate il secondo giorno= q_{um}(martedì)

25 =15+10 domande cumulate il terzo giorno= q_{cum}(mercoledi)

37= 25+ 12 domande cumulate il quarto giorno= q_{cum}(giovedì)

60= 37+13 domande cumulate il quinto giorno= q_{cum}(venerdi)

q_{cum}: $g \in \{lun., mar., mer., gio., ven.\} \rightarrow f_{cum}(g)$ = *n.* **domande** ***presentate entro g***

Il carattere è ordinato e a ogni sua modalità puo' essere assegnato un codice che rappresenti l'ordine in cui vengono considerate le modalità

1= lunedì
2= martedì
3= Mercoledì
4= Giovedì
5= venerdi

X = val.	$q_a(x)$	n. dei casi con determinazione del carat. ≤x
1	7	7
2	8	15
3	10	25
4	12	37
5	23	60
totale	60	

7=$q_{cum}(1)$

15= $q_{cum}(2)$

25 =$q_{cum}(3)$

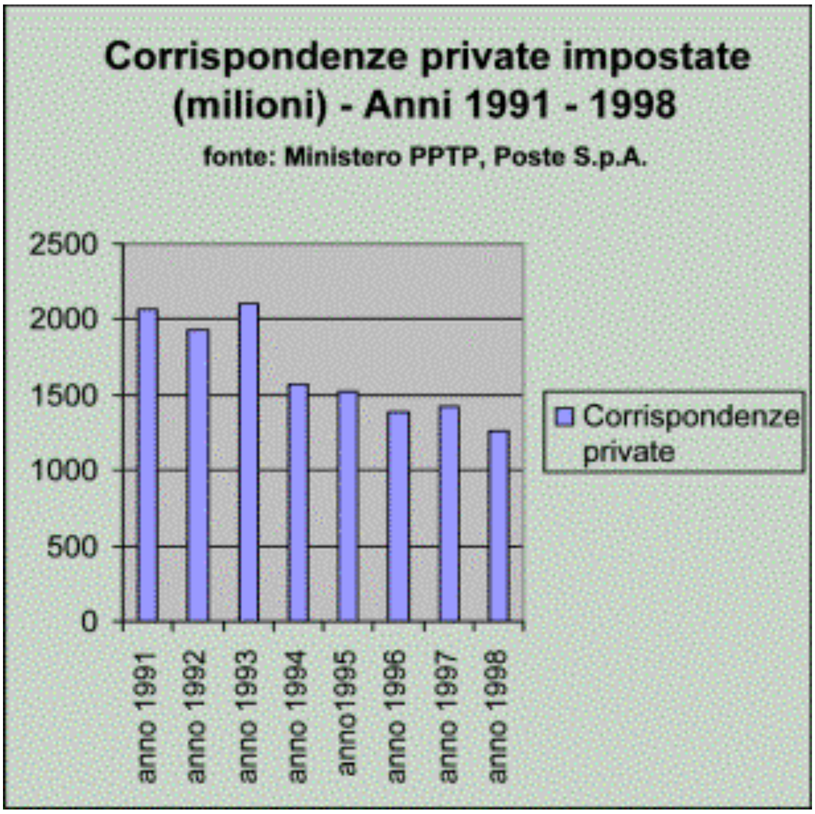
37= $q_{cum}(4)$

60= $q_{cum}(5)$ =totale

$q_{cum}: x \in \{1, 2, 3, 4, 5\} \rightarrow f_{cum}(x)=$ ***n. casi con detrminazione del carattere ≤ x***

funzione strettamente crescente $x_1 < x_2 \Rightarrow q_{cum}(x_1) < q_{cum}(x_2)$

Esempio di ortogramma rappresentante una serie di quantità e lettura delle quantità cumulate attraverso l'ortogramma.



Come si legge l'ortogramma

1. **il *titolo*: corrispondenze private impostate (lettere, cartoline, pacchi…) in Lombardia negli anni dal 1991 al 1998. Il dato è espresso in miliardi di copie.**

2. **la *fonte*: Ministero PPTP, Poste SpA (cioè gli enti che si occupano della corrispondenza)**

3. **la *legenda*: alla base di ogni rettangolo, la cui base per convenzione misura 1, è registrato l'anno di rilevamento dei dati; l'altezza di ogni rettangolo, pari all'area dello stesso, è invece misurabile tramite una scala graduata (qui varia tra zero e 2000 milioni, come è precisato nel titolo). Ad esempio nel 1991 sono stati inoltrati poco più di 2 miliardi di buste e affini, mentre nel 1998 non si arriva agli 1,3 miliardi. Questo dato è il più basso della serie.**

4. **Volendo sapere quanti miliardi di lettere sono state inviate tramite le poste italiane negli anni dal 1991 al 1998 non si fa altro che sommare tutti i dati, cioè le altezze dei rettangoli (siamo nell'ordine di grandezza dei 13 miliardi!); in particolare se si vuole la quantità di corrispondenza impostata entro l'anno 1994 (cioè la quantità cumulata all'anno 1994), occorre sommare le altezze dei primi quattro rettangoli.**

4. Dalla distribuzione semplice di frequenze cumulate alla distribuzione semplice di frequenze assolute

Siano $x_1, x_2, \ldots, x_k$ le modalità dal carattere e risulti

$$x_1 < x_2 < \ldots < x_k .$$

Allora

$f_{cum}(x_1) = f_a(x_1)$

$f_{cum}(x_2) = f_{cum}(x_1) + f_a(x_2)$

$f_{cum}(x_3) = f_{cum}(x_2) + f_a(x_3)$

……

$f_{cum}(x_k) = f_{cum}(x_{k-1}) + f_a(x_k) = n$ ***Tot. unità stat.***

Dalle suindicate uguaglianze otteniamo $n = f_{cum}(x_k)$ ***e le seguenti uguaglianze***

$f_a(x_1) = f_{cum}(x_1)$

$f_a(x_2) = f_{cum}(x_2) - f_{cum}(x_1)$

$f_a(x_3) = f_{cum}(x_3) - f_{cum}(x_2)$

……

$f_a(x_k) = f_{cum}(x_k) - f_{cum}(x_{k-1})$

che permettono di risalire dalla distribuzione semplice delle frequenze cumulate assolute alla distribuzione semplice delle frequenze assolute

Esempio

Distribuzione statistica relativa alla variabile ordinata " condizione dei mezzi di trasporto"

Modalità	$f_{cum}(m)$	$\rightarrow$	$f_a(m)$
Scadenti	***32***	$\rightarrow$	***32***
Accettabili	***89***	$\rightarrow$	***89-32=57***
Efficienti	***98***	$\rightarrow$	***98-89=9***
Tot. casi	***98***		

5 Appendice: due diverse definizioni di frequenza cumulata

$f_{cum}(x)$ = n. unità statistiche che presentano il carattere in una determinazione $\leq x$

valori x_i	$f_a\,(x_i)$	**f_a cum. ≤** **numero dei casi con valore ≤ x_i**
1	**7**	**7 (casi con valore 1 o minore di 1)**
2	**10**	**17 (casi con valore 2 o minore di 2)**
3	**8**	**25 (casi con valore 3 o minore di 3)**
4	**3**	**28 (casi con valore 4 o minore di 4)**
5	**2**	**30 (casi con valore 5 o minore di 5)**
	30	

$f^{\circ}_{cum}(x)$ = n. unità statistiche che presentano il carattere in una determinazione < x

valori x_i	$f_a\,(x_i)$	**$f^{\circ}_{cum}(x)$ =** **numero dei casi con valore < x_i**
1	**7**	**0 (casi con valore minore di 1)**
2	**10**	**7 (casi con valore minore di 2)**
3	**8**	**17 (casi con valore minore di 3)**
4	**3**	**25 (casi con valore minore di 4)**
5	**2**	**28 (casi con valore minore di 5)**
(6)		**30 (casi con valore minore di 6)**
	30	

Per un carattere quantitativo con valori raggruppati in classi chiuse inferiormente è conveniente riferisi alla seconda definizione di frequenza cumulata

Esercizio: disegnare le ogive delle due distribuzioni di frequenze cumulate

Esercizio

TABELLE DEI DATI GREZZI RELATIVI AD UN'INDAGINE SULLE FAMIGLIE
DISTRIBUZIONE UNITARIA MULTIPLA

FAMIGLIA	NUMERO DI COMPONENTI	REDDITO ANNUO (in milioni)	TITOLO DI STUDIO DEL CAPOFAMIGLIA	ZONA DI RESIDENZA
1	2	28	Elementare	Nord
2	1	35	Media Inferiore	Centro
3	3	50	Media Inferiore	Nord
4	1	45	Media Superiore	Nord
5	1	40	Laurea	Sud (isole)
6	2	30	Media Inferiore	Sud (isole)
7	3	55	Media Inferiore	Centro
8	4	80	Media Superiore	Centro
9	5	60	Laurea	Sud (isole)
10	6	85	Laurea	Nord
11	7	90	Laurea	Nord
12	1	52	Media Superiore	Centro
13	2	62	Media Superiore	Sud (isole)
14	3	75	Media Superiore	Sud (isole)
15	5	60	Elementare	Nord
16	4	45	Media Inferiore	Nord
17	3	42	Media Inferiore	Centro
18	2	28	Elementare	Nord
19	8	70	Media Superiore	Sud (isole)
20	2	38	Laurea	Sud (isole)

La tabella è relativa alla rilevazione di quattro caratteri su di un campione di 20 famiglie
- Determinare per ogni carattere una tabella una tabella modalità/ frequenza (assoluta, relativa e percentuale) e rappresentare graficamente le diverse distribuzioni.
- Per i caratteri per cui è possibile farlo indicare anche le distribuzioni di frequenze cumulate e rappresentarle graficamente.
Infine
- Considerare il carattere reddito come un carattere continuo, suddividere i valori in classi, e dare rappresentazione sia della distribuzioni di frequenza assoluta, che di quella cumulata.