

Tre organizzazioni di un reparto di produzione

- *Frequenze assolute*
- *Diagramma a bastoncini*
- *Funzione di ripartizione empirica*
- *Distribuzione di frequenza in classi*
- *Istogramma*
- *Frequenze relative e cumulate*

Il problema

In un reparto dove sono assemblati *riproduttori MP3* vengono, in tre giorni diversi, provate tre differenti organizzazioni delle linee di produzione. Le tre diverse organizzazioni sono chiamate nel seguito

- *vecchia* (quella in uso al momento dell'esperimento),
- *nuova_1*,
- *nuova_2*.

Nei tre giorni, per ciascuno dei 288 addetti che lavorano nel reparto, viene rilevato il numero di operazioni completate, che, ovviamente, può essere visto come una misura della produttività.

Domanda:
qual è l'organizzazione del lavoro migliore?

I dati

Vecchia organizzazione

725 724 710 724 700 724 713 692 683 712 684 707 703 691 709 702 705 715
704 705 697 725 692 719 694 717 696 707 726 703 705 712 710 697 698 694
701 715 701 707 706 701 687 708 719 713 699 702 694 708 712 704 703 687
709 693 715 707 710 700 718 702 718 705 723 718 701 698 692 684 716 710
708 707 695 726 710 709 692 707 717 709 710 718 708 720 705 714 687 707
707 723 695 676 705 684 717 719 715 710 711 696 696 715 686 702 708 713
701 692 713 700 704 726 702 706 706 700 700 687 696 694 699 709 704 704
715 706 688 724 713 686 697 710 704 724 721 717 690 707 713 685 706 699
687 702 701 708 704 705 702 701 699 699 685 712 678 706 706 695 707 718
706 716 703 721 714 704 697 693 711 697 710 713 702 715 714 716 698 714
704 717 700 692 718 699 698 690 710 703 702 719 710 725 721 713 699 703
698 712 714 707 691 711 712 718 702 711 709 700 719 692 716 700 707 714
717 714 703 709 711 704 689 712 714 711 692 720 697 698 700 689 693 707
699 704 696 708 713 714 712 708 704 720 705 703 712 719 713 716 712 703
717 695 711 697 693 701 699 697 724 713 706 705 704 707 704 719 711 700
694 706 705 698 697 697 700 705 722 712 703 688 694 708 703 690 706 704

Organizzazione ``nuova 1''

695 686 694 690 713 704 693 697 723 694 690 721 683 701 718 715 738 694
692 704 728 697 711 706 714 710 717 729 709 695 699 714 691 698 680 720
683 696 713 674 689 683 708 704 725 695 690 696 678 725 683 700 699 705
688 714 709 693 681 717 691 706 684 684 693 719 731 706 686 698 710 679
712 688 697 729 695 697 717 679 736 671 695 739 698 696 714 711 701 720
686 706 722 695 688 709 693 756 677 712 670 693 695 683 713 672 706 708
690 685 686 681 716 709 704 679 686 676 718 683 689 696 687 736 699 685
698 700 723 681 713 700 708 705 718 692 743 715 745 700 693 676 723 712
671 714 687 687 687 683 671 677 696 696 714 713 671 688 675 671 692 725
690 680 693 703 733 708 720 704 688 732 711 685 714 704 686 682 699 708
708 704 685 685 694 702 738 702 696 709 701 687 703 701 702 693 691 701
735 721 705 691 741 685 716 716 737 687 732 697 670 684 693 711 685 705
690 705 693 698 678 704 710 686 689 686 698 684 687 696 719 679 696 701
712 691 686 704 744 705 718 709 725 699 721 690 678 713 714 705 681 721
673 698 717 711 670 726 694 723 701 683 716 671 712 704 699 705 727 719
702 692 708 694 670 694 697 682 718 705 699 709 695 711 688 717 699 686

organizzazione ``nuova 2''

698 715 675 710 731 721 705 718 693 702 713 730 707 710 744 725 724 701
737 715 704 723 705 702 698 729 698 723 716 698 732 724 721 722 728 740
727 709 724 746 704 740 729 708 721 714 739 713 752 732 713 692 734 727
725 690 749 706 758 722 697 722 705 723 748 730 706 688 709 739 709 744
704 716 748 713 744 721 723 733 707 723 702 734 690 715 711 705 718 702
706 742 742 736 740 712 722 731 713 704 704 735 700 717 746 735 717 718
691 696 720 735 716 745 714 698 709 704 704 684 749 747 715 717 731 700
747 709 705 749 704 697 694 715 737 734 705 726 710 716 740 731 714 733
726 752 714 710 714 753 749 728 696 733 731 728 686 706 710 729 729 730
722 707 716 702 728 716 743 750 715 735 710 734 712 706 719 709 702 712
710 729 728 720 721 752 715 712 717 692 724 720 739 719 712 713 734 734
710 711 722 743 707 729 712 681 739 699 721 706 703 708 719 708 724 730
726 731 734 739 727 759 718 716 715 719 693 729 738 710 730 726 719 726
733 717 701 723 720 744 730 698 729 696 717 713 705 700 715 710 735 726
732 701 707 724 708 730 721 720 706 700 735 706 725 725 735 695 709 705
702 737 688 727 717 708 720 724 731 706 730 714 703 721 712 748 734 724

Organizzazione dei dati in una distribuzione di frequenza

I dati non sono “tantissimi” rispetto ad altre situazioni che si possono incontrare nelle applicazioni. Sono però troppi per cercare di capire qualcosa solamente “guardandoli”. Dobbiamo quindi cercare di sintetizzarli in qualche modo.

Il carattere osservato (numero di osservazioni completate) è “discreto” e assume solo valori interi. Possiamo in prima battuta individuare i valori minimo (670) e massimo (759), elencare tutti i possibili valori compresi tra questi due e contare quanti operai hanno completato esattamente un dato numero di operazioni in ciascuno dei tre giorni.

Si ottengono in questo modo tre **distribuzioni di frequenza assoluta** (riportate a pagina successiva con il nome vecchia, nuova_1 e nuova_2)

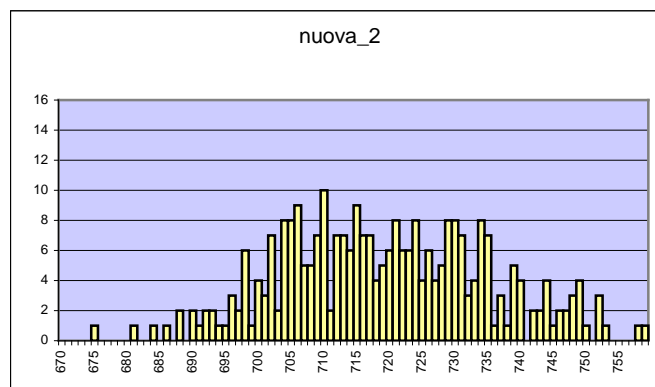
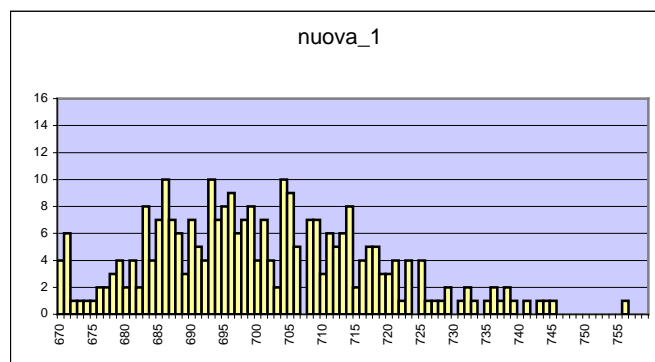
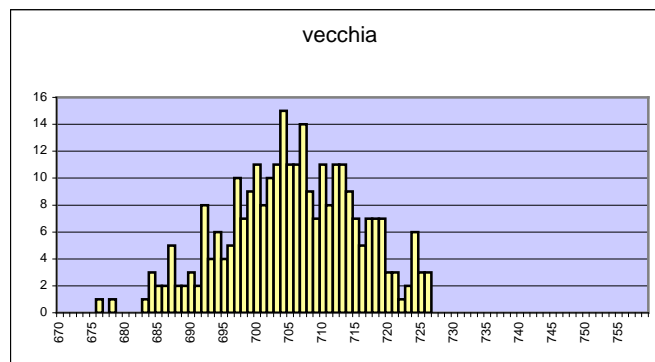
In generale, il termine **frequenza assoluta** indica *il numero di unità statistiche che soddisfano una certa condizione* (nel nostro caso, che presentano la modalità indicata nella prima colonna).

n. operaz completate	vecchia	nuova_1	nuova_2
670		4	
671		6	
672		1	
673		1	
674		1	
675		1	1
676	1	2	
677		2	
678	1	3	
679		4	
680		2	
681		4	1
682		2	
683	1	8	
684	3	4	1
685	2	7	
686	2	10	1
687	5	7	
688	2	6	2
689	2	3	
690	3	7	2
691	2	5	1
692	8	4	2
693	4	10	2
694	6	7	1
695	4	8	1
696	5	9	3
697	10	6	2
698	7	7	6
699	9	8	1
700	11	4	4
701	8	7	3
702	10	4	7
703	11	2	2
704	15	10	8
705	11	9	8
706	11	5	9
707	14		5
708	9	7	5
709	7	7	7
710	11	3	10
711	8	6	2
712	11	5	7
713	11	6	7
714	9	8	6
715	7	2	9

n. operaz completate	vecchia	nuova_1	nuova_2
716	5	4	7
717	7	5	7
718	7	5	4
719	7	3	5
720	3	3	6
721	3	4	8
722	1	1	6
723	2	4	6
724	6		8
725	3	4	4
726	3	1	6
727		1	4
728		1	5
729		2	8
730			8
731		1	7
732		2	3
733		1	4
734			8
735		1	7
736		2	1
737		1	3
738		2	1
739		1	5
740			4
741		1	
742			2
743		1	2
744		1	4
745		1	1
746			2
747			2
748			3
749			4
750			1
751			
752			3
753			1
754			
755			
756		1	
757			
758			1
759			1
totale	288	288	288

Le modalità diverse sono ancora molte e non è agevole trarre indicazioni dall'esame delle tre distribuzioni.

Un giudizio più immediato lo si può trarre dai **diagrammi a bastoncini**, ottenuti disegnando in corrispondenza di ogni valore osservato un bastoncino (perpendicolare all'asse delle x) di lunghezza uguale alla frequenza assoluta con cui quel valore è stato osservato.

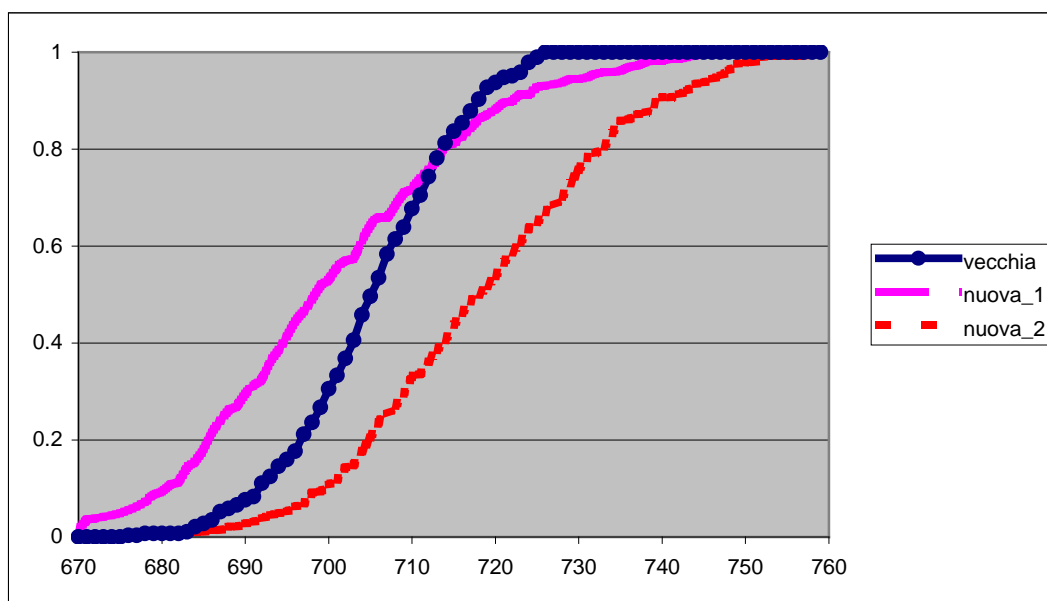


Funzione di ripartizione empirica

Un'altra rappresentazione grafica di uso frequente è offerta dalla **funzione di ripartizione empirica** (che, tra l'altro, ha altre importanti applicazioni) così definita.

$$\boxed{\begin{array}{l} \text{funzione di} \\ \text{ripartizione} \\ \text{empirica} \\ \text{calcolata in} \end{array}} = \frac{\boxed{\begin{array}{l} \text{numero di} \\ \text{osservazioni} \\ \text{minori di } x \end{array}}}{\boxed{\begin{array}{l} \text{numero totale} \\ \text{di} \\ \text{osservazioni} \end{array}}}$$

Per le tre organizzazioni del lavoro, il grafico di queste funzioni è il seguente.



Il messaggio può forse sembrare a prima vista meno evidente di quello contenuto nei diagrammi a bastoncini. Ma se si guarda alla definizione, ci si convince che il messaggio è il medesimo.

Raggruppamento in classi di modalità

Nel nostro caso, l'elencazione di tutte le modalità diverse appare dispersiva (difficile cogliere un differente livello di produttività tra chi ha completato, ad esempio, 700 o 701 operazioni).

Una alternativa consiste nel suddividere l'intervallo che contiene tutti i valori osservati ([670 - 759]) in un certo numero di “sotto-intervalli” e poi semplicemente nel “contare” quante osservazioni cadono nei vari sotto-intervalli.

Utilizzando sotto-intervalli di lunghezza 5 ed aperti a destra, si ottiene la tabella seguente. (distribuzione di frequenza con modalità raggruppate in classi)

operazioni	vecchia	nuova 1	nuova 2
[670 - 675)	0	13	0
[675 - 680)	2	12	1
[680 - 685)	4	20	2
[685 - 690)	13	33	3
[690 - 695)	23	33	8
[695 - 700)	35	38	13
[700 - 705)	55	27	24
[705 - 710)	52	28	34
[710 - 715)	50	28	32
[715 - 720)	33	19	32
[720 - 725)	15	12	34
[725 - 730)	6	9	27
[730 - 735)	0	4	30
[735 - 740)	0	7	17
[740 - 745)	0	3	12
[745 - 750)	0	1	12
[750 - 755)	0	0	5
[755 - 760)	0	1	2
totale	288	288	288

Nota alla tabella: La prima colonna mostra i sotto-intervalli utilizzati. Le altre mostrano il numero di addetti che hanno completato un numero di operazioni appartenenti al sottointervallo considerato. Ad esempio, il 13 che compare nella prima riga alla colonna 3 indica che esattamente 13 dei 288 addetti hanno, nel giorno in cui è stata sperimentata l'organizzazione nuova_1, completato un numero di operazioni maggiore od uguale di 670 e minore (strettamente) di 675.

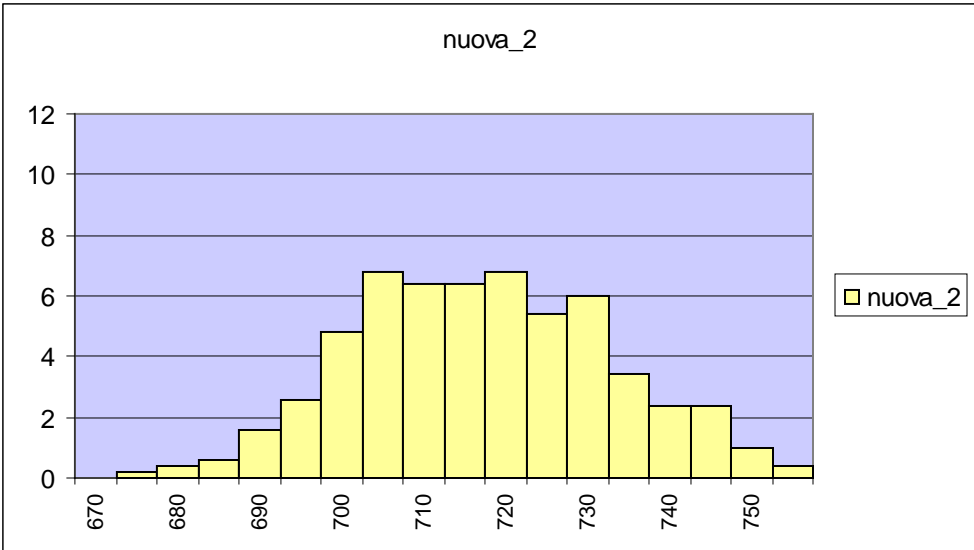
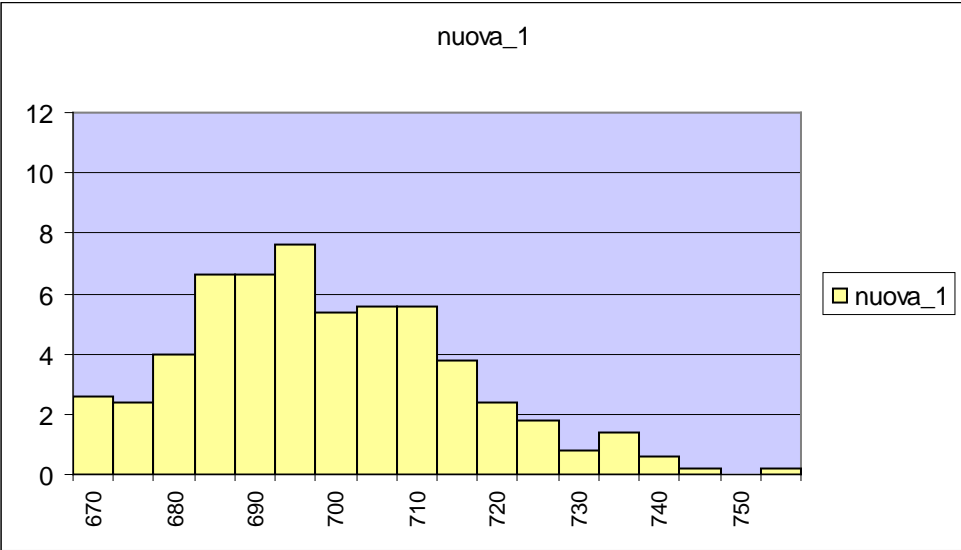
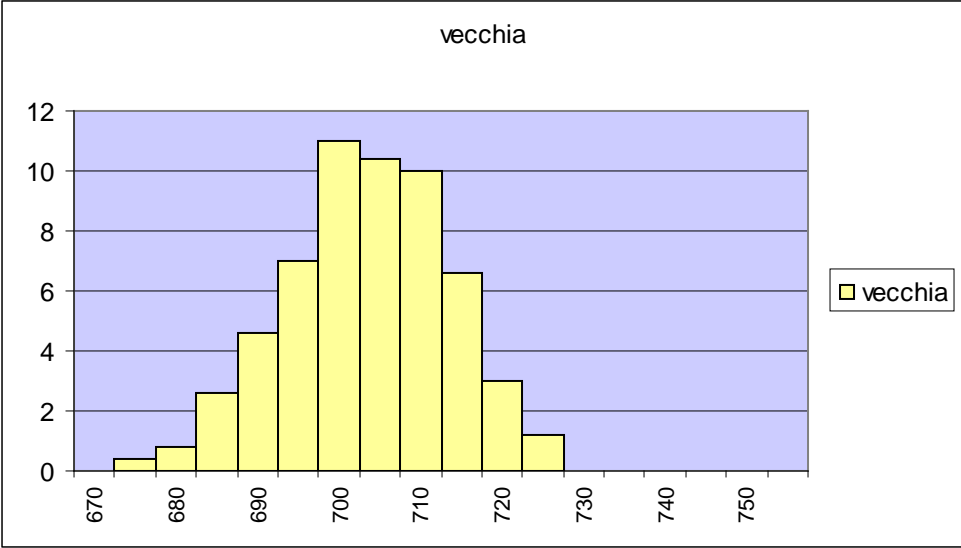
Commento alla tabella: nuova_2 sembra l'organizzazione migliore; nuova_1 è probabilmente l'organizzazione peggiore.

Istogramma

Le differenze tra le tre distribuzioni di frequenza si colgono ancora meglio se le rappresentiamo graficamente. Una possibilità è nella pagine seguente. Il grafico è stato costruito ponendo

(base rettangoli) = (intervallini riportati nella 1a colonna)
(altezza rettangoli) = (frequenze/base)

I diagrammi del tipo mostrato sono chiamati **istogrammi**.



Commento al grafico: Il grafico suggerisce le stesse considerazioni fatte sulla base della tabella.

La distribuzione di nuova_2 è, rispetto alle altre, quella più spostata verso destra (ovvero verso livelli di maggiore produttività). Nuova_2 appare quindi come l'organizzazione migliore (sulla base di questi dati).

Frequenze relative

Dividendo ciascuna frequenza assoluta per il numero totale di unità statistiche (288 nel nostro caso) otteniamo le cosiddette **frequenze relative**, che hanno il vantaggio, rispetto alle frequenze assolute, di permettere di confrontare distribuzioni di frequenza basate su numeri differenti di unità statistiche.

operazioni completate	% vecchia	% nuova_1	% nuova_2
[670 - 675)	0.0	4.5	0.0
[675 - 680)	0.7	4.2	0.3
[680 - 685)	1.4	6.9	0.7
[685 - 690)	4.5	11.5	1.0
[690 - 695)	8.0	11.5	2.8
[695 - 700)	12.2	13.2	4.5
[700 - 705)	19.1	9.4	8.3
[705 - 710)	18.1	9.7	11.8
[710 - 715)	17.4	9.7	11.1
[715 - 720)	11.5	6.6	11.1
[720 - 725)	5.2	4.2	11.8
[725 - 730)	2.1	3.1	9.4
[730 - 735)	0.0	1.4	10.4
[735 - 740)	0.0	2.4	5.9
[740 - 745)	0.0	1.0	4.2
[745 - 750)	0.0	0.3	4.2
[750 - 755)	0.0	0.0	1.7
[755 - 760)	0.0	0.3	0.7
totale	100.0	100.0	100.0

Frequenze cumulate

Si ottengono “cumulando” progressivamente le frequenze. Sono essenzialmente analoghe alla funzione di ripartizione empirica.

Possono essere “assolute” o “relative”. Quelle relative coincidono con la funzione di ripartizione empirica alla fine di ogni intervallo.

Esempio di calcolo (nuova 1)

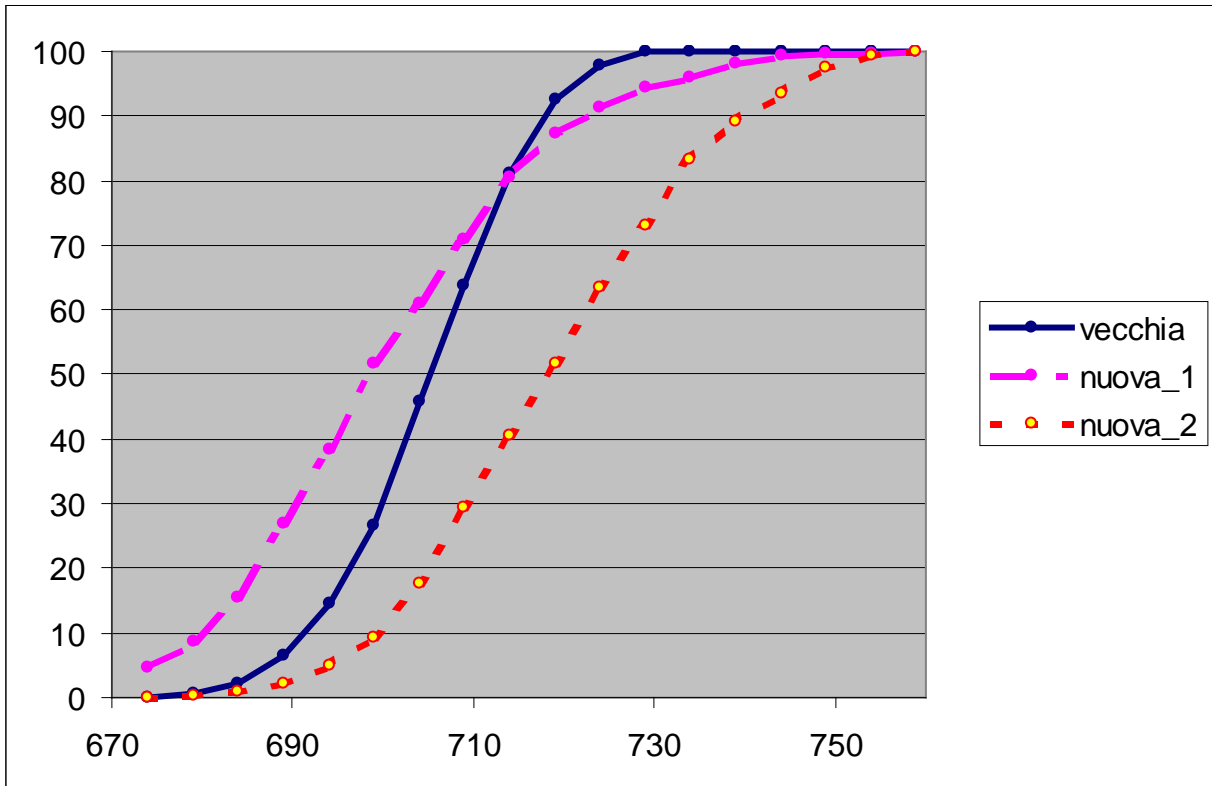
fine intervallo	freq	operazione	cumulata
675	13	13	13
680	12	13+12	25
685	20	20+13+12	45
690	33	33+20+13+12	78

.....

Frequenze cumulate assolute e relative (%)

operazioni complete	Freq cumulate assolute			Freq cumulate %		
	vecchia	nuova_1	nuova_2	vecchia	nuova_1	nuova_2
[670 - 675)	0	13	0	0.0	4.5	0.0
[675 - 680)	2	25	1	0.7	8.7	0.3
[680 - 685)	6	45	3	2.1	15.6	1.0
[685 - 690)	19	78	6	6.6	27.1	2.1
[690 - 695)	42	111	14	14.6	38.5	4.9
[695 - 700)	77	149	27	26.7	51.7	9.4
[700 - 705)	132	176	51	45.8	61.1	17.7
[705 - 710)	184	204	85	63.9	70.8	29.5
[710 - 715)	234	232	117	81.3	80.6	40.6
[715 - 720)	267	251	149	92.7	87.2	51.7
[720 - 725)	282	263	183	97.9	91.3	63.5
[725 - 730)	288	272	210	100.0	94.4	72.9
[730 - 735)	288	276	240	100.0	95.8	83.3
[735 - 740)	288	283	257	100.0	98.3	89.2
[740 - 745)	288	286	269	100.0	99.3	93.4
[745 - 750)	288	287	281	100.0	99.7	97.6
[750 - 755)	288	287	286	100.0	99.7	99.3
[755 - 760)	288	288	288	100.0	100.0	100.0
totale	288	288	288	100.0	100.0	100.0

Funzione di ripartizione empirica alla fine di ogni intervallo (cumulata relativa)



Misure di posizione

- Media aritmetica
- Mediana
- Quantili, quartili e percentili
- Diagramma a scatola con baffi

Misure o parametri di posizione

Le distribuzioni dei pezzi prodotti differiscono, come visto, soprattutto per la diversa “posizione”.

Una domanda che sembra naturale è “di quanto?”.

Ad esempio, nuova_2 sembra, con i dati a disposizione, migliore di vecchia. Ma quanto migliore?

Una possibile maniera per rispondere a questo tipo di domande si concretizza nel

1. Sintetizzare le singole distribuzioni in un unico numero che, in una qualche senso, indichi dove la distribuzione stessa è “posizionata”. Ovvero, calcolare per ogni distribuzione una **misura (o parametro o indice) di posizione**.
2. Rispondere confrontando gli indici calcolati al punto precedente.

Famosi parametri di posizione sono: la **media aritmetica**, la **mediana** e i **quantili**.

La media aritmetica

Supponiamo di aver rilevato un certo fenomeno (esprimibile numericamente) su n unità statistiche diverse. Indichiamo con y_1, y_2, \dots, y_n i valori osservati (ovvero, i nostri dati). La media aritmetica dei dati è

$$\bar{y} = \frac{y_1 + y_2 + \dots + y_{n-1} + y_n}{n} = \frac{1}{n} \cdot \sum_{i=1}^n y_i$$

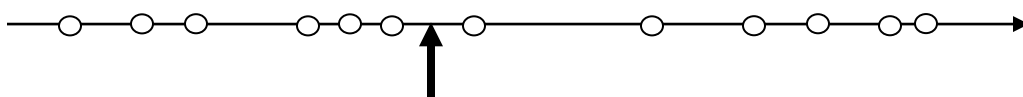
Esistono molti tipi di “medie”. Quella aritmetica è senza ogni dubbio quella di utilizzo più comune. Per questo motivo, viene comunemente indicata come “la media” senza nessuna ulteriore aggettivazione.

La mediana

E' la modalità osservata nella unità statistica che occupa la posizione centrale della graduatoria ordinata in senso crescente.

- Se le unità sono in numero dispari, è la modalità osservata nell'unità che occupa il posto $(n+1)/2$.
- Se le unità sono in numero pari, è un valore intermedio alle modalità osservate nelle unità che occupano le posizioni $n/2$ e $(n+1)/2$.

Ad esempio, nel grafico seguente, supponendo che le osservazioni corrispondano ai punti disegnati con una 'o', un possibile valore per la mediana è stato indicato con una freccia. Infatti, il punto così marcato lascia sia a sinistra che a destra 6 osservazioni.



Media e mediana: il caso delle tre organizzazioni del lavoro

posizione in graduatoria	vecchia	nuova_1	nuova_2
1	676	670	675
2	678	670	681
3	683	670	684
4	684	670	686
5	684	671	688
.....
142	705	699	718
143	705	699	718
144	706	699	718
145	706	699	719
146	706	699	719
147	706	699	719
.....
284	725	741	752
285	725	743	752
286	726	744	753
287	726	745	758
288	726	756	759
totale	203176	201828	207123
n	288	288	288
Media	705.5	700.8	719.2
Mediana	706.0	699.0	718.5

Come si vede i risultati confermano le impressioni precedenti. Indicano che nuova 2 potrebbe far aumentare la produzione di circa un 2%.

Quantili

Generalizzano la mediana.

L'idea alla base di un **quantile-p** dove $p \in [0; 1]$ è di cercare un numero che sia più grande del $100 \times p\%$ dei dati osservati e più piccolo del restante $100 \times (1 - p)\%$. Ad esempio, un quantile-0.1 deve essere un valore che lascia a sinistra il 10% delle osservazioni ed a destra il restante 90%.

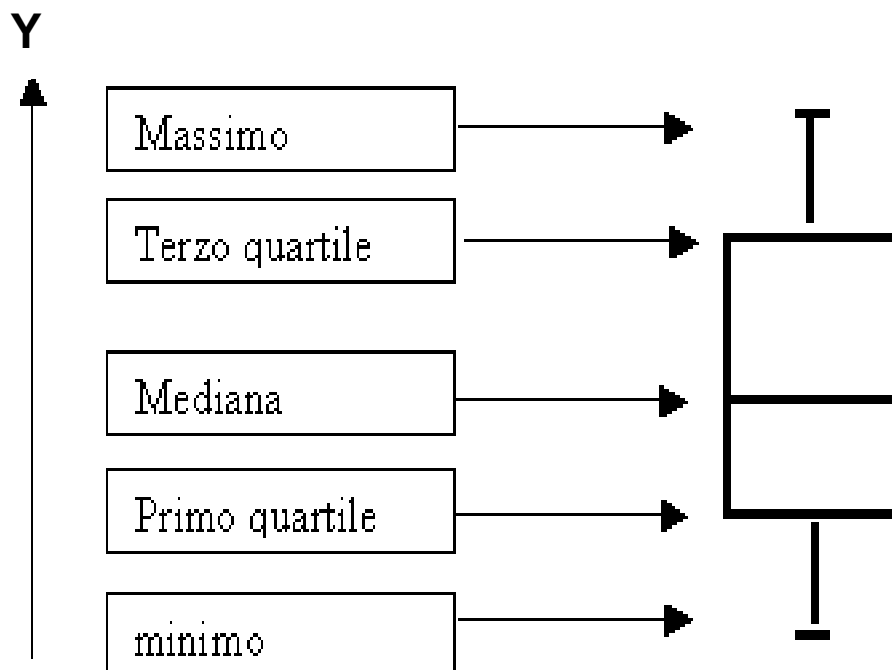
Si osservi che, per costruzione, il quantile-p è il valore y_i per il quale è $F(y_i) \cong p$ dove con $F(\bullet)$ abbiamo indicato la funzione di ripartizione empirica.

I quantili con p uguale a 0;25, 0;50 e 0;75 vengono chiamati rispettivamente il primo, il secondo e il terzo **quartile**. Dividono la popolazione in quattro parti uguali. Si osservi che il 2° quartile coincide con la mediana.

I quantili con $p = 0; 0.1; \dots ; 0.99$ si chiamano **percentili**.

Diagrammi a scatola con baffi

- Il nome deriva dall'inglese (*box and whiskers plot* spesso, anche in italiano, abbreviato in *boxplot*).
- Forniscono una idea schematica di un insieme di dati basata sui quantili.
- Sono costituiti, come dice il nome, da una *scatola* e da due *baffi* costruiti in accordo al disegno sottostante.



Tre organizzazioni della produzione: boxplot

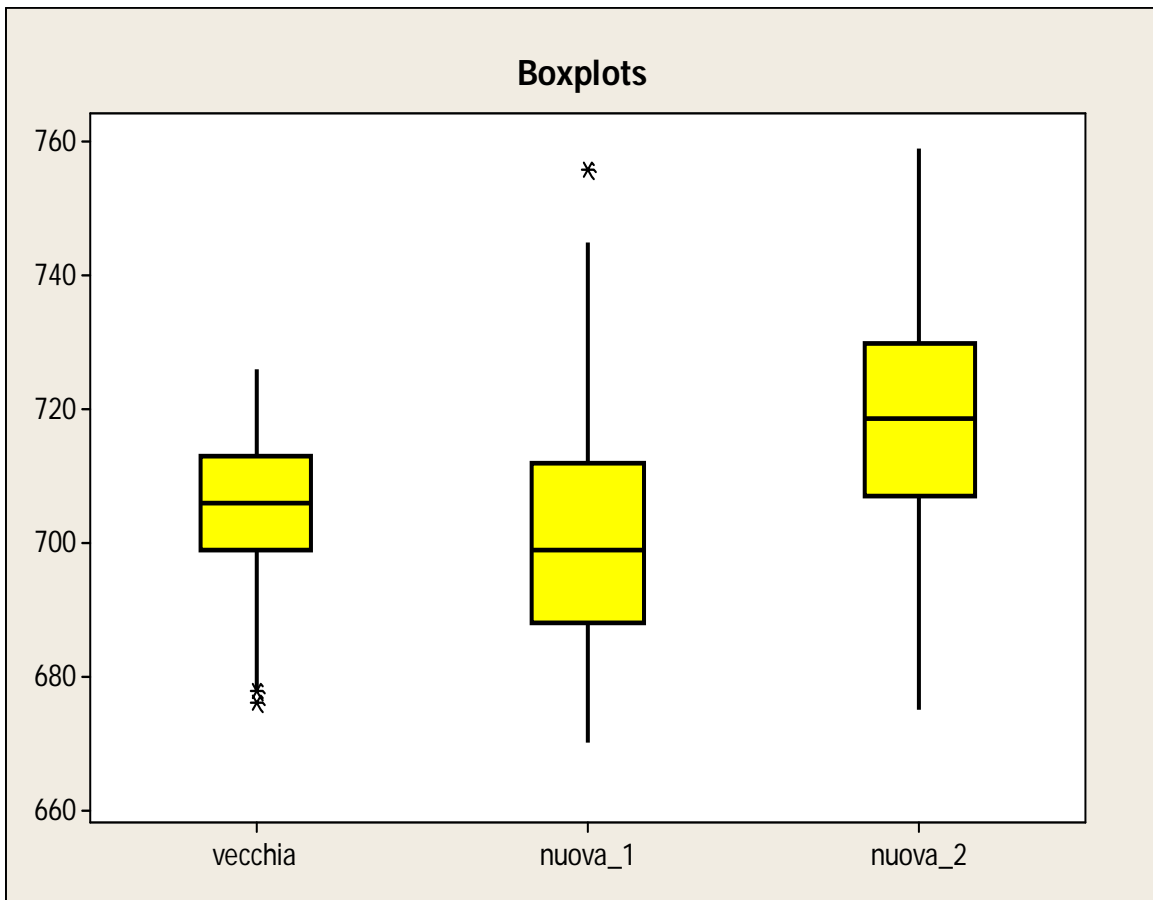
	posizione in graduatoria	vecchia	nuova_1	nuova_2
Minimo	1	676	670	675
	2	678	670	681

1° quartile	72	699	688	707
1° quartile	73	699	688	707

Mediana	144	706	699	718
Mediana	145	706	699	719

3° quartile	216	713	712	730
3° quartile	217	713	712	730

	287	726	745	758
Massimo	288	726	756	759



Tre organizzazioni della produzione: misure di variabilità e scomposizione della devianza

CONSIDERANDO LE TRE GIORNATE SEPARATAMENTE

	vecchia	nuova_1	nuova_2
media	705.5	700.8	719.2
Devianza	28661.78	79241.50	71043.97
Varianza	99.52	275.14	246.68
SQM	9.98	16.59	15.71
CV(*100)	1.41	2.37	2.18

CONSIDERANDO LE TRE GIORNATE COMPLESSIVAMENTE

N	864	288*3
Media totale	708.48	
Devianza Totale	231531.67	
Varianza	267.98	
SQM	16.37	

Calcolo della devianza tra i gruppi (tra le medie dei gruppi)	Mj	M	(Mj-M)^2	Nj	(Mj-M)^2*Nj
	705.4722222	708.4803241	9.048676751	288	2606.0189
	700.7916667	708.4803241	59.11545273	288	17025.25
	719.1770833	708.4803241	114.4206587	288	32953.15
					52584.419

Devianza vecchia	28661.78	
Devianza nuova_1	79241.50	
Devianza nuova_2	71043.97	
Devianza ENTRO	178947.25	77.3%
Devianza TRA	52584.42	22.7%
Devianza Totale	231531.67	100.0%