Elementi essenziali di item analisi e statistiche descrittive

di Giorgio Asquini e *Luca Piria*

5.1 PREMESSA*

Perché in un corso dedicato alla formazione di un tecnico della valutazione dedicare un intero modulo (20 ore, il 20% del tempo del corso) ad aspetti matematici e statistici?

Perché siamo convinti, con Tullio De Mauro, che "fare seriamente didattica comporta sapersi fare i conti" (De Mauro, *Prolegomeni ad una ridefinizione della didattica*, in "Università Progetto", 6, 1985). Nel brano citato De Mauro non si riferisce ad un uso tipicamente umanista dell'espressione "fare i conti", inteso cioè come "confronto con il passato" o "analisi dei pro e dei contro". Intende proprio addizioni e divisioni, calcoli insomma.

Occuparsi di formazione professionale rientra a pieno titolo in una operazione didattica (cioè insegnamento e apprendimento), e una componente fondamentale della didattica è la valutazione, per cui un tecnico di valutazione degli apprendimenti non può ignorare i supporti matematici che una scienza importante come la statistica può offrire a chi vuole impostare correttamente molte delle procedure didattiche connesse alla valutazione.

Nel corso di questo capitolo si cercherà sempre di motivare contestualmente le ragioni del ricorso all'analisi statistica. Qui cerchiamo di fornire una serie di motivazioni quadro, cercando anche di anticipare alcune facili obiezioni.

Lo sviluppo della statistica è legato alla consapevolezza della variabilità dei dati: se tutti i fenomeni (da quelli fisici a quelli pedagogici) si realizzassero in modo identico, non ci sarebbe diversità negli esiti, che sarebbero sempre predeterminati. A che servirebbe una scienza che studia la variabilità dei fenomeni e ragiona in ter-

* L'ideazione del seguente capitolo è frutto di un lavoro comune, tuttavia Giorgio Asquini è autore dei parr. 5.1, 5.2, 5.3, 5.5, 5.6, 5.7, 5.10, Luca Piria dei parr 5.4, 5.8, 5.9, e ha curato gli Esercizi e l'Allegato.

mini di probabilità o possibilità degli esiti? Ma il meccanicismo non è certo una corrente di pensiero che ha avuto successo nel nostro secolo (basti pensare alla relatività e alla meccanica quantistica), e lo svolgimento dei fenomeni si è rivelato un po' più complicato rispetto ai rassicuranti meccanismi di causa-effetto della fisica tradizionale.

Anche per le scienze umanistiche si è passati da un determinismo ferreo ad una relativizzazione complessa: dalla ricerca di principi unificanti che spiegassero in modo analogico tutti i fenomeni legati all'uomo (dalla psicologia, alla sociologia, alla pedagogia) si è passati all'analisi dei molteplici fattori che possono influenzare anche i fenomeni studiati dalle scienze umane, fattori che tendono a combinarsi fra loro in modo sempre diverso, o meglio in molti modi diversi.

La consapevolezza della complicazione dei fenomeni ha prodotto i diversi incroci fra le scienze umane (ma è stata anche da questi alimentata), per cui oggi non ha più senso parlare di una scienza-pedagogia autonoma rispetto alla scienza-psicologia. Accanto a queste contaminazioni è cresciuta l'esigenza di utilizzare strumenti più idonei per la lettura dei fenomeni, per poter considerare una quantità sempre maggiore di eventi da un gran numero di punti di vista. La strumentazione migliore è quella fornita dall'analisi statistica, che ha ricevuto un grosso impulso proprio dai problemi posti, fin dall'inizio del Novecento, dalla psicologia sociale, alla ricerca di nuovi metodi di analisi applicabili a gruppi consistenti di persone.

Lo scopo principale dell'analisi statistica è quindi una lettura più precisa dei fenomeni, non fondata solo sulle impressioni dell'osservatore. Anche chi si occupa di formazione deve operare delle scelte basandosi su una serie di esperienze, e la statistica serve proprio a descrivere con maggiore precisione le esperienze e a fornire possibilità di scelta a chi deve decidere.

Una frequente obiezione avanzata dagli insegnanti è che le informazioni ottenibili dall'analisi statistica sono quasi sempre ricavabili da fonti più immediate, sintetizzabili nella capacità degli insegnanti di "leggere" i fenomeni che si verificano nella classe. E' indubbio che un insegnante abile e navigato sia in grado di anticipare le informazioni fornite da qualsiasi analisi di tipo statistico applicata al suo gruppo di allievi, ma facciamo notare che:

- non tutti gli insegnanti sono abili e navigati, soprattutto all'inizio della loro carriera, per cui rischiano di non cogliere, o ancora peggio di travisare, il significato dei fenomeni che si verificano sotto i loro occhi;

- anche gli insegnanti abili possono basare la loro lettura delle situazioni su un'esperienza poco variata, per cui tendono a dare significati stereotipati alle situazioni, imputando al caso (o a entità esterne) i fatti che non riescono a spiegare;
- gli insegnanti esperti spesso sono in difficoltà nel descrivere a colleghi o altre persone coinvolte nel processo didattico la loro lettura della situazione del gruppo di allievi, proprio perchè basata su elementi ritenuti affidabili (l'esperienza) ma indefinibili ("Vedrai, andranno bene all'esame, me lo sento").

L'analisi statistica non intende certo soppiantare l'esperienza o il "sesto senso" dell'insegnante, ma semplicemente fornirgli più stimoli e possibilità di verifica: se anche vengono confermate completamente le intuizioni dell'insegnante, non è forse un ottimo segnale (per la sua professionalità) e un incoraggiamento (per proseguire l'azione didattica)? E se anche le informazioni provenienti dall'analisi non coincidono con le sensazioni dell'insegnante, non è forse utile avviare una verifica (dei materiali, del proprio operato, del lavoro degli allievi ...) prima che si determinino inconvenienti difficili da risolvere?

Un'altra obiezione mossa all'analisi statistica applicata alla valutazione è la sua difficoltà materiale di applicazione, che si concretizza in procedure lente e a rischio di errore. Insomma il tempo impiegato (per alcuni perso) non varrebbe i risultati ottenibili.

Il fondamento di questa obiezione è facilmente verificabile nel diffuso imbarazzo riscontrabile, in qualsiasi gruppo di persone, quando si comincia a parlare in termini matematici: non pochi alzano bandiera bianca, confessando una naturale idiosincrasia con numeri e operazioni (per fare questa affermazione ci fidiamo della pratica e non ci fondiamo su dati statistici). Ma il motivo di questa incompatibilità è una caratteristica tipicamente italiana, in particolare legata alla scarsa efficacia dell'insegnamento della matematica nella scuola superiore, e questa non è una sensazione, ma il risultato dei confronti internazionali (analisi statistiche) svolti dalle organizzazioni che studiano l'efficacia dei sistemi educativi, ed in particolare la IEA (International Association for the Evaluation of Educational Achievement, il cui referente italiano è il Centro Europeo Dell'Educazione - CEDE di Frascati) e l'OECD-OCSE (Organisation for Economic Co-operation and Development – Organizzazione per lo Sviluppo e la Cooperazione Economica).

Abbiamo cercato di trasformare il limite in una condizione di partenza: tutte le procedure presentate in questo capitolo verranno ampiamente illustrate anche nei dettagli, senza dare per scontate conoscenze pregresse in campo statistico, anzi fornendo esempi e proponendo esercizi per verificare l'applicabilità delle operazioni indicate. Cercheremo anche di sfruttare una possibilità sempre più disponibile per chiunque voglia o debba utilizzare l'analisi statistica: il supporto informatico. E' in continua crescita la diffusione di programmi, in particolare fogli elettronici, che facilitano notevolmente l'analisi statistica, riducendo tempi di calcolo o possibilità di errore, e aumentando le possibilità di elaborazione e rappresentazione dei dati.

Nello svolgimento del capitolo verranno proposte soprattutto procedure manuali, o semi-automatizzate, per insistere sulla comprensione della procedura, ma è importante sapere che, una volta padroneggiati i significati e le tecniche, molte operazioni delle procedure statistiche possono essere svolte in modo automatico.

Nella prima parte del capitolo presentiamo le modalità di realizzazione dell'**item analisi**, procedura utile per la verifica (come strumento) delle prove oggettive. Nella seconda parte verranno invece illustrate le procedure di realizzazione delle **statistiche descrittive**, cioè le misurazioni relative al gruppo di allievi che svolge una prova oggettiva, procedure utili per giungere ad una valutazione delle abilità verificate con le prove. Le due procedure sono strettamente collegate. Infatti come guida operativa utilizzeremo i dati provenienti da un unico gruppo-esempio (19 soggetti), impegnato in una prova-esempio (10 quesiti a scelta multipla), di cui non forniremo il testo, proprio per concentrare l'attenzione sul trattamento dei dati.

5.2 COME SI COSTRUISCE UNA TABELLA DI DATI

Per un'analisi corretta di dati è necessario seguire scrupolosamente una serie di procedure. Una delle più delicate è sicuramente la tabulazione delle risposte. In questo paragrafo cercheremo proprio di affrontare e risolvere i problemi di tabulazione, per poter successivamente passare alle modalità di analisi vere e proprie dei dati tabulati. Tabulare significa fare tabelle. Fino a pochi anni fa questa operazione si svolgeva esclusivamente a mano, con riga e matita, su un foglio quadrettato; nella tabella così disegnata si trascrivevano poi i dati, cioè le risposte al test. Un'impostazione errata della tabella, o la necessità di modificarne gli elementi, costringeva spesso al suo rifacimento completo. Ecco perché una corretta procedura di creazione della tabella permetteva di risparmiare molto tempo. Oggi i computer, e in particolare i programmi del tipo foglio elettronico, permettono di tabulare i dati molto più velocemente, e di modificare agevolmente gli elementi della tabella; ma la semplificazione di eventuali modifiche non ha cambiato i criteri di funzionalità della tabulazione, per cui i dati devono essere inseriti anche nel computer con determinati accorgimenti.

Nel nostro caso - valutazione di prove oggettive costituite da un numero limitato di quesiti e applicate in gruppi non troppo estesi - non possiamo considerare scontato l'uso del computer, per cui illustreremo la tabulazione e l'analisi manuale, o meglio assistita solo dalla calcolatrice tascabile, dei dati ottenuti dall'applicazione delle prove oggettive. Sarà semplice, per chi abitualmente utilizza il computer, trasferire le indicazioni qui fornite per svolgere tabulazione e analisi informatica dei dati. Nella bibliografia sono comunque presenti dei testi anche con indicazioni specifiche per i programmi più diffusi, come Excel o Works.

Per impostare correttamente una tabella è necessario riflettere preventivamente sui dati disponibili e su quelli che si vogliono ottenere. Per riferirci a un esempio specifico simuliamo lo svolgimento di una prova di comprensione della lettura composta da 10 quesiti a scelta multipla (con quattro alternative di risposta) in una classe di 19 allievi. La soluzione più immediata sembrerebbe essere la costruzione di una tabella di 10 colonne per 19 righe, ma non è così perché:
- servono una colonna e una riga in più per le etichette (dei quesiti e degli alunni);
- serve una colonna o una riga in più per le chiavi di risposta (per ogni quesito);
- serve una riga o una colonna in più per i totali delle risposte corrette (per allievo);
- servono delle colonne o delle righe in più (per ogni quesito) per riportare i totali di tutte le alternative di risposta;
- servono delle colonne o delle righe in più (per ogni quesito) per riportare gli indici calcolabili sulle risposte date dagli allievi.

Il dubbio relativo a colonne o righe in più dipende dall'orientamento che vogliamo dare alla tabella: nell'esempio che qui proponiamo abbiamo deciso che gli studenti occupino le colonne, mentre i quesiti le righe, per cui la tabella che dobbiamo costruire sarà così composta:

Colonne	Righe
1 per le etichette dei quesiti	1 per le etichette di allievi e di altre colonne
19 per gli allievi	10 per i quesiti
1 per le chiavi (risposte corrette dei quesiti)	1 per i totali delle risposte corrette (allievi)
4 per i totali di ogni alternativa	
2 per gli errori di risposta e le omissioni	
2 per gli indici calcolabili	
Totale 29 colonne	Totale 12 righe

Quindi dobbiamo preparare una tabella di 29 colonne per 12 righe.

Quanto deve essere larga ogni colonna e alta ogni riga? La risposta è strettamente collegata alle dimensioni del foglio che utilizziamo, ma per rendere la tabella più leggibile possiamo dire che:

- riga e colonna delle etichette devono essere abbastanza ampie da contenere informazioni leggibili, per esempio se definisco gli allievi con i loro cognomi devo poterli scrivere per esteso;

- le celle (cioè le caselle formate dall'incrocio di una riga con una colonna) che contengono le risposte possono essere strette, poiché dovrò trascrivere solo una lettera o un numero;

- le celle che contengono i totali devono invece essere abbastanza ampie per contenere fino a 6-7 cifre o segni (se uso il formato percentuale e i decimali).

Il risultato di queste impostazioni è costituito dalla tabella della fig.5.1

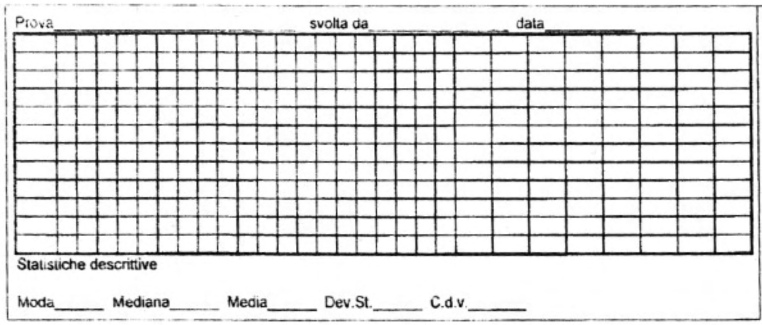

Fig. 5.1 - Tabella modello per la tabulazione dei dati

Nella figura 5.1, nella parte bassa, sono stati predisposti degli spazi per trascrivere alcune misure utili per l'analisi descrittiva dei

dati, in particolare: moda, mediana, media, deviazione standard, coefficiente di variazione. Torneremo più avanti sul significato e sull'uso di queste misure. Prima di passare all'operazione di tabulazione dei dati dobbiamo anche ricordare che nel nostro esempio consideriamo una prova formata da elementi omogenei (tutti quesiti a scelta multipla con quattro distrattori). Può succedere però di tabulare prove formate da elementi diversi (quesiti vero/falso, scelte multiple con diverso numero di alternative, corrispondenze, completamenti etc.); in questo caso la tabulazione deve prevedere queste diversità per due aspetti fondamentali: la pesatura dei singoli punteggi e lo spazio necessario nella tabella per l'item analisi.

Circa la pesatura dei punteggi rimandiamo al capitolo dedicato alla costruzione degli strumenti in cui è stato già spiegato come distinguere i punteggi grezzi secondo il tipo di item. Ricordiamo qui solamente che per svolgere l'item analisi bisogna ordinare i punteggi degli allievi, per cui è necessario predeterminare il peso di ogni risposta corretta: nel nostro esempio (quesiti dello stesso tipo) ogni risposta corretta vale 1, ma è chiaro che in una prova mista la risposta corretta di un quesito vero/falso non può avere lo stesso peso rispetto ad uno a scelta multipla.

Per lo spazio necessario nella tabella bisogna considerare il numero massimo di alternative possibili, anche per un solo quesito, e prevedere nella tabella un numero adeguato di colonne per le alternative di risposta. Se la prova prevede alcuni quesiti con meno alternative, nelle righe corrispondenti verranno compilate solo le prime colonne, in numero adeguato per le due (nei quesiti vero/falso) o più alternative previste.

5.3 LA TABULAZIONE DEI DATI

Una volta costruita la tabella modello, su carta o su computer, si può passare alla tabulazione dei dati. Per svolgere questa operazione è opportuno lavorare su una copia del modello, mantenendo la tabella originale non compilata per successive tabulazioni.

La tabulazione dei dati deve essere necessariamente preceduta dalle correzione delle prove. Lavorando con il computer anche questa operazione può essere automatizzata, ma per comprendere il meccanismo è meglio seguire i diversi passi della procedura manuale.

Usando una penna o matita rossa si corregge una prova alla volta (l'ordine è ininfluente in questa fase); una codifica semplice può consistere nel cerchiare il numero d'ordine di ogni domanda per cui è stata scelta la risposta corretta; in tal modo il numero dei cerchi corrisponde al punteggio ottenuto dall'allievo, che si può trascrivere nella parte alta della prova. Ma è necessario anche segnalare gli errori e le omissioni, per cui si possono sottolineare le risposte errate, scrivere una piccola "E" accanto alle lettere dei distrattori in caso di risposta non data correttamente (per esempio tracciando due crocette), scrivere una piccola "O" sotto le domande senza risposta.

Una volta corrette le prove bisogna ordinarle per punteggio, dal più basso al più alto; in caso di punteggi uguali l'ordine interno è ininfluente. Ora è tutto pronto per la trascrizione dei dati sulla tabella. Si comincia trascrivendo nella seconda colonna a destra (la prima è riservata alle etichette degli item) i dati dello studente che ha ottenuto il punteggio più basso. Nella prima riga bisogna trascrivere il suo nome o il suo codice (il numero d'ordine del registro per esempio), nella seconda la risposta alla prima domanda, cioé l'alternativa scelta ("A", "B", "C" o "D"), con l'accortezza di non trascrivere nulla in caso di risposta corretta. In questo modo gli spazi bianchi corrisponderanno al punteggio ottenuto dall'allievo, che deve essere trascritto nell'ultima riga, quella del totale. Naturalmente bisogna tabulare anche gli errori di compilazione (una "e" minuscola per vederli meglio dentro la tabella) e le omissioni (una "o").

Prova ______ svolta da ______ data ______

	7
item1	B
item2	C
item3	C
item4	A
item5	
item6	B
item7	C
item8	D
item9	B
item10	o
	1

Statistiche descrittive

Moda_____ Mediana_____ Media_____ Dev.St._____ C.d.v._____

Fig. 5.2 - Tabulazione dei dati del primo studente (n.7)

Terminato il primo studente si passa al secondo seguendo la stessa procedura. Proseguendo nella trascrizione diminuiranno sempre più i dati da trascrivere, e se alcuni studenti hanno dato tutte le risposte giuste le loro colonne saranno vuote. Questo accorgimento oltre al risparmio di compilazione permette una prima stima visiva dei risultati della prova, ricavabile dall'equilibrio di pieni e vuoti della tabella.

Prova ______ svolta da ______ data ______

	7	5	12	9	17	6	19	11	3	1	15	18	4	13	2	10	14	8	16
item1	B	D	D		B	B	A			A	B	B			D				
item2	C	C		B					C										
item3	C	B	A	C	B	B	B	C	B	B	C	B		B			B		
item4	A	A						D	D	A			A	B	B	B			
item5																			
item6	B	D	A	A	D	B	D	A					B						
item7	C		B		C					B	B								
item8	D	B	D	B	A	B	B	B	B	A	A	B	A	B	D	A	B	B	
item9	B	C	C	B		B		C	D				C						
item10	o			o	o		o					o							
	1	3	4	4	4	5	5	5	5	5	6	6	6	7	7	8	8	9	10

Statistiche descrittive

Moda_____ Mediana_____ Media_____ Dev.St._____ C.d.v._____

Fig. 5.3 - Tabulazione dei dati di tutti gli studenti

Terminata questa prima fase di trascrizione si possono riporre le prove e passare al completamento della tabella. Si può scegliere di completare subito le colonne della parte destra (totali e indici) o le misure per l'analisi descrittiva, nella parte bassa. Si tratta di procedure autonome, per cui l'ordine di compilazione è indifferente. Nel nostro esempio completiamo prima la tabella, cioè procediamo con l'item analisi.

5.4 LA PROCEDURA DI ITEM ANALISI

L'item analisi è una procedura per verificare l'efficacia di una prova, e in particolare il funzionamento delle domande, delle risposte corrette e dei distrattori. Nelle ricerche svolte su un grande campione questa procedura serve a validare strumenti per ampie rilevazioni, per cui si basa su un grande numero di risposte. Nella pratica didattica quotidiana un controllo sull'efficacia degli item può fornire diverse indicazioni, anche se la prova è stata già svolta da altri gruppi di allievi ed è già stata sottoposta ad analisi. Procedendo nella presentazione della procedura vedremo quali indicazioni si possono ricavare da un item analisi effettuata su un piccolo numero di allievi.

Per prima cosa contiamo le risposte corrette (spazi vuoti) di ogni item, riportando i risultati nella colonna con l'etichetta corrispondente alla chiave di risposta, indicata nella colonna specifica.

Poi contiamo per ogni item le scelte raccolte da ogni distrattore e trascriviamo i relativi totali nelle colonne specifiche. Un esempio per chiarire la procedura si può vedere nella figura 5.4: in questa ipotesi l'alternativa corretta (chiave di risposta) è la "C", con 9 scelte (spazi vuoti), poi ci sono 2 "A", 5 "B", 3 "D", nessun errore e nessuna omissione.

	7	5	12	9	17	6	19	11	3	1	15	18	4	13	2	10	14	8	16	Ch	A	B	C	D	e	o		
Item1	B	D	D		B	B	A			A	B	B			D					C	2	5	9	3	0	0		

Fig. 5.4 - Trascrizione dei totali risposta corretta e distrattori

E' possibile, per diversi aspetti consigliabile, riportare i totali direttamente in formato percentuale, ottenibile con la seguente operazione:

$$\text{Valore percentuale} = \frac{\text{totale risposte alternativa}}{\text{numero degli alunni}} * 100$$

In tal caso la tabella risulterà compilata come la figura 5.5, in cui sono stati effettuati arrotondamenti al primo decimale.

	7	5	12	9	17	6	19	11	3	1	15	18	4	13	2	10	14	8	16	Ch	A	B	C	D	e	o		
item1	B	D	D		B	B	A			A	B	B			D					C	10,5	26,3	47,4	15,8	0	0		

Fig. 5.5 - Trascrizione dei totali risposta corretta e distrattori (con valori percentuali)

Si può passare ora alla parte più importante della procedura di item analisi, il calcolo degli indici.

Cominciamo con l'**indice di facilità**, che ci permette di verificare il livello di difficoltà del quesito. In pratica si tratta di mettere in relazione gli studenti che hanno risposto correttamente con il numero complessivo degli studenti. Tanti più studenti hanno risposto correttamente alla domanda, tanto più questa risulterà facile. Ecco la formula per calcolare l'indice:

$$\text{Indice di facilità} = \frac{\text{totale risposte corrette}}{\text{numero degli alunni}}$$

L'indice risultante può variare da un minimo di 0 (nessuna risposta corretta, massima difficoltà) a un massimo di 1 (tutte risposte corrette, massima facilità). In alcuni manuali questo indice è chiamato "indice di difficoltà", ma é preferibile adottare la denominazione "indice di facilità" in quanto al crescere del punteggio (da 0 a 1) aumenta la facilità del quesito, non la sua difficoltà.

Tornando al nostro esempio ecco il calcolo dell'indice applicato all'Item1:

$$\text{Indice di facilità Item1} = \frac{9}{19} = 0{,}47$$

Come si può facilmente osservare il risultato della formula (0,47) é la semplificazione del calcolo percentuale riferito all'alternativa corretta (47,4%). L'indice così ottenuto va trascritto nella tabella, e in particolare nella colonna etichettata "Fac" (vedi Fig. 5.6). In genere vengono considerati poco efficaci gli item con un indice di difficoltà inferiore a 0,25, in quanto troppo difficili, o superiore a 0,75, perché troppo facili. Come vedremo però per valutare l'efficacia dell'item questo dato deve essere affiancato dall'indice di discriminatività, per cui i limiti segnalati sono solo indicativi.

	7	5	12	9	17	6	19	11	3	1	15	18	4	13	2	10	14	8	16	Ch	A	B	C	D	e	o	Fac	
Item1	B	D	D		B	B	A			A	B	B			D					C	10,5	26,3	47,4	15,8	0	0	0,47	

Fig. 5.6 - Trascrizione dell'indice di facilità

Il calcolo dell'**indice di discriminatività** risulta essere leggermente più complesso. Da un punto di vista strettamente matematico l'operazione ha un nome preciso, cioè "correlazione punto-biseriale", ed é un'operazione che un foglio di calcolo elettronico svolge agevolmente basandosi proprio sui dati delle risposte degli studenti. Prima di presentare la formula illustriamo i criteri di costruzione dell'indice. Per discriminatività si intende la capacità di un item di distinguere gli studenti più abili dagli studenti meno abili. I primi sono quelli che hanno ottenuto i punteggi più alti nell'intera prova (composta da tutti gli item), mentre i secondi sono gli studenti con i punteggi più bassi. L'indice di discriminatività quindi confronta le risposte degli studenti nel singolo item, ma considerando il risultato complessivo della prova. Nella tabulazione delle risposte abbiamo usato l'accorgimento di ordinare gli studenti da sinistra (punteggi più bassi) a destra (punteggi più alti). Bisogna considerare due gruppi di studenti (detti "estremi") della stessa consistenza, corrispondenti ognuno a circa 1/3 dell'intero gruppo (restano quindi esclusi gli studenti della fascia centrale). Chiamiamo estremo superiore il gruppo di studenti con i punteggi più alti, estremo inferiore il gruppo con i punteggi più bassi. Fatte queste puntualizzazioni vediamo la formula:

$$\textbf{Indice di discriminatività} = \frac{\text{risp.esatte Estremo superiore - risp.esatte Estremo inferiore}}{\text{numero degli studenti di un estremo}}$$

Ricordiamo sempre che l'indice va calcolato per ogni item, sottraendo ogni volta le risposte corrette dei meno abili (estremo inferiore) dalle risposte corrette dei migliori (estremo superiore), dividendo infine l'esito della sottrazione per il numero di studenti dell'estremo. Anche in questo caso analizziamo i limiti di variazione dell'indice. Se le risposte corrette degli studenti più abili sono in quantità maggiore (é il caso più frequente) l'indice sarà positivo e misurerà la discriminatività dell'item, raggiungendo il valore massimo di 1 quando avremo tutte risposte corrette dell'estremo superiore contro nessuna risposta corretta dell'estremo inferiore; più raramente può succedere che l'estremo inferiore risponda più correttamente del superiore, in tal caso l'indice sarà negativo, e verrà misurata la discriminatività negativa, cioè l'item rovescia il criterio di abilità che la prova complessiva intende misurare, con un minimo teorico di -1 nel caso in cui tutti i migliori sbaglino e tutti i peggiori rispondano correttamente. Se le risposte corrette dei due estremi si eguagliano (non importa la quantità effettiva) l'indice risulta 0, che corrisponde alla mancanza di discriminatività, cioè quell'item non riesce a distinguere l'abilità misurata dalla prova. Nella pratica vengono considerati efficaci gli item che hanno un indice di discriminatività superiore a 0,30, mentre devono essere sicuramente scartati gli item con un indice negativo; gli item con indice positivo ma vicino allo zero possono essere recuperati se l'indice di facilità é valido, con eventuali modifiche in edizioni successive della prova per renderli più discriminativi.

Come abbiamo fatto per l'indice di facilità esemplifichiamo il calcolo della discriminatività per l'item1:

$$\text{Indice di discriminatività Item1} = \frac{5 - 1}{6} = 0{,}66$$

Come si può vedere dalla Fig. 5.7 (sono stati evidenziati con un fondino i due estremi) tra i migliori 6 studenti ci sono state 5 risposte corrette (spazi vuoti), contro 1 sola risposta corretta dei 6 studenti meno abili: la quantità dei due gruppi (6 studenti) corrisponde a un terzo circa dell'intero gruppo (19 studenti). Nel nostro esempio l'Item1 presenta un buon livello di discriminatività (0,66 trascritto nella colonna con etichetta "Dis"), che unito a un buon indice di facilità (0,47) attesta l'efficacia complessiva dell'item.

	7	5	12	9	17	6	19	11	3	1	15	18	4	13	2	10	14	8	16	Ch	A	B	C	D	e	o	Fac	Dis
Item1	B	D	D		B	B	A			A	B	B			D					C	10,5	26,3	47,4	15,8	0	0	0,47	0,66

Fig. 5.7 - Trascrizione dell'indice di discriminatività

5.5 VALIDAZIONE DI UN ITEM

Abbiamo visto come ogni quesito sia definibile con due indici, facilità e discriminatività, e abbiamo applicato, a titolo esemplificativo, il calcolo di questi indici all'Item1 della nostra prova di esempio. Proseguiamo nel calcolo degli indici per tutti gli altri item della prova, cercando di definirne la validità, cioè l'efficacia per il tipo di abilità che la prova intende misurare. Applichiamo pertanto le formule sopra indicate ai 9 quesiti rimanenti, e subito dopo analizziamo, per ogni item i risultati ottenuti.

Prova__________ svolta da__________ data________

	7	5	12	9	17	6	19	11	3	1	15	18	4	13	2	10	14	8	16	Ch	A	B	C	D	e	o	Fac	Dis
item1	B	D	D		B	B	A			A	B	B			D					C	10,5	26,3	47,4	15,8	0	0	0,47	0,66
item2	C	C		B					C											A	78,9	5,3	15,8	0	0	0	0.79	0,50
item3	C	B	A	C	B	B	B	C	B	B	C	B		B			B			D	5,3	47,4	21,0	26,3	0	0	0,26	0,66
item4	A	A						D	D	A			A	B	B	B				C	21,0	15,8	52,7	10,5	0	0	0,53	-0,16
item5																				B	0	100	0	0	0	0	1	0
item6	B	D	A	A	D	B	D	A					B							C	15,8	15,8	52,6	15,8	0	0	0,53	1
item7	C		B		C					B	B									D	0	15,8	10,5	73,7	0	0	0,74	0,50
item8	D	B	D	B	A	B	B	B	B	A	A	B	A	B	D	A	B	B		C	26,3	52,6	5,3	15,8	0	0	0,05	0,16
item9	B	C	C	B		B		C	D				C							A	57,9	15,8	21,0	5,3	0	0	0,58	0,83
item10	o			o	o		o					o								D	0	0	0	73,7	0	26,3	0,74	0,50
	1	3	4	4	4	5	5	5	5	5	6	6	6	7	7	8	8	9	10									

Statistiche descrittive

Moda_____ Mediana_____ Media_____ Dev.St._____ C.d.v.______

Fig. 5.8 – Trascrizione degli indici per tutti gli item

L'item2 risulta estremamente semplice (0,79 di facilità), ma adeguatamente discriminativo (0,50). Se analizziamo i distrattori vediamo che nessuno ha scelto l'alternativa D, che pertanto dovrà essere modificata se si intende salvare l'item. Anche il distrattore B risulta debole (una sola scelta), e dovrebbe essere modificato. Il salvataggio di un item dubbio, come questo deve essere deciso considerando il complesso della prova: se risulta l'unico item critico può essere abbandonato senza problemi, se invece ci sono item con indici peggiori (questo risulta almeno discriminativo) conviene tentare il recupero, rendendo più plausibili i distrattori B e D. Si può però anche considerare l'ipotesi di conservare il quesito tale e quale, però facendolo diventare il primo della prova: proporre agli allievi inizialmente un quesito facile può risultare motivante verso la prova, per cui si può accettare anche un indice di facilità elevato, sempre che gli altri item presentino una facilità accettabile.

L'item3 risulta abbastanza difficile (0,26), ma con una buona discriminatività. Il problema in questo caso riguarda il distrattore B, che raccoglie un numero molto alto di scelte (quasi la metà), superiore alla risposta corretta, per cui deve essere rivisto, ed eventualmente modificato per renderlo meno distraente, a favore della risposta corretta e del distrattore A, che è stato scelto da un solo allievo.

L'item4 presenta un caso abbastanza inconsueto: accanto ad un indice di facilità più che accettabile (0,53) troviamo una discriminatività negativa (-0,16), cioè a questa domanda hanno risposto più correttamente gli allievi meno abili (4 risposte giuste) rispetto ai compagni migliori (3 risposte giuste). Anche visivamente si può notare che le risposte errate si addensano maggiormente sulla destra della tabella. Evidentemente il quesito non è costruito in maniera corretta, ed induce maggiormente in errore gli studenti più abili. Da rivedere con attenzione il distrattore B, scelto, guarda caso, solo dagli studenti dell'estremo superiore. Evidentemente si tratta di un distrattore troppo sofisticato, che mette in difficoltà solo gli studenti più abili, probabilmente perché non risulta del tutto errato anche rispetto alla risposta corretta C. Questo tipo di item deve essere eliminato se non si può intervenire sul distrattore mal funzionante.

L'item5 è facilmente analizzabile: è l'unico che ha ricevuto tutte risposte corrette, quindi il massimo della facilità e il minimo della discriminatività. Di fatto risulta inutile per la valutazione e può essere escluso dalla prova. Questo caso estremo si verifica quando l'item è riferito ad una conoscenza o a un'abilità che sono già padroneggiate con sicurezza da tutto il gruppo che svolge la

prova. In realtà, se la prova fosse stata costruita correttamente, un caso del genere non dovrebbe presentarsi, poichè il redattore della prova dovrebbe essere in grado di prevedere in anticipo il livello di abilità del gruppo di allievi. Ma utilizzando prove già svolte in altri gruppi si potrebbe riscontrare un'eccessiva facilità di alcuni quesiti per il nuovo gruppo di allievi. Il confronto permette proprio di scoprire su quali argomenti ci sia una diversità fra i gruppi. In ogni caso il quesito però non deve essere buttato, perchè potrebbe rivelarsi adeguato in una prova costruita per un livello di abilità inferiore, dove l'argomento dell'item risulterebbe discriminante.

L'item6 risulta perfettamente adeguato, con una buona facilità (0,53) e una discriminatività massima (1). Da notare anche il perfetto equilibrio dei distrattori, che raccolgono tutti la stessa percentuale di scelte.

L'item7 risulta agli estremi di accettabilità per l'indice di facilità (0,74), ma con una discreta discriminatività (0,50). Sicuramente da migliorare il distrattore A, che non ha ricevuto scelte, rivelandosi quindi poco plausibile per il gruppo di allievi che ha svolto la prova.

L'item8 presenta indici di facilità e discriminatività molto bassi (rispettivamente 0,05 e 0,16), perché solo uno studente risponde correttamente. L'item pertanto risulta inutilizzabile, perché non adeguato, per la sua forte difficoltà, al livello di abilità del gruppo. Da notare che più di metà degli allievi ha indicato il distrattore B, che è risultato fin troppo plausibile. Una attenta analisi del contenuto di questo distrattore può fornire utili indizi per capire i limiti degli allievi circa l'argomento del quesito, anche per attuare un successivo intervento didattico chiarificatore. Anche in questo caso, come per l'item troppo facile, il quesito può essere conservato per una prova di livello superiore.

L'item9 presenta un andamento molto simile al 6, con una buona facilità (0,58) e alta discriminatività (0,83). Leggermente debole il distrattore D, ma i buoni indici permettono di considerare il quesito utilizzabile anche senza modifiche.

Infine l'item10 ha indici uguali all'item7 (0,74 di facilità e 0,50 di discriminatività), ma i tre distrattori non sono stati scelti e ci sono ben cinque omissioni di risposta. In pratica il quesito risulta molto facile, per gli studenti che hanno risposto, e possiamo ipotizzare (la verifica va fatta con le osservazioni provenienti da chi ha somministrato la prova) che gli studenti che non hanno risposto non sono riusciti a considerare il quesito per mancanza di tempo. Le omissioni si presentano infatti spesso sui quesiti finali: se riguardano uno o due studenti possono essere legate a insicurezza personale, ma se, come nel nostro esempio, riguardano circa un quarto del gruppo è probabile che il tempo assegnato sia stato troppo limitato. L'analisi di questo item deve pertanto essere sospesa è rimandata ad un'ulteriore somministrazione. Se consideriamo infatti solo le risposte date, l'indice di facilità è 1 (troppo facile), e la discriminatività non riguarda l'abilità che la prova intende misurare, ma l'uso del tempo, abilità certo importante ma non confondibile con la comprensione o il profitto.

La nostra prova esempio non poteva esaurire tutti i casi possibili, per cui bisogna aggiungere alcune ipotesi di lettura. Se le omissioni si verificano su un item iniziale o intermedio vuol dire che gli allievi lo hanno saltato e non ci sono più ritornati. In questo caso la lettura degli indici (e le scelte per i distrattori) devono cercare di indagare i motivi dell'omissione; per esempio se l'item risulta discriminativo per le omissioni dell'estremo inferiore, potrebbe significare che i meno bravi non sono riusciti a decidere quale fosse la risposta più plausibile, per cui l'item risulterebbe funzionante, distrattori compresi. Se le omissioni si ritrovano invece soprattutto nell'estremo superiore (discriminatività bassa o negativa), è possibile che ci sia un distrattore troppo plausibile (probabilmente corretto), per cui gli studenti più abili, nel dubbio e rispettando la consegna di indicare una sola risposta, si siano astenuti; in tal caso il quesito deve essere eliminato o profondamente rivisto.

In genere, se le istruzioni sono state chiare, è difficile rilevare errori nelle modalità di risposta (infatti nel nostro esempio la relativa colonna è sempre a 0). Quando si verificano, gli errori tendono a concentrarsi per un allievo o per un item. In entrambi i casi si tratta di una cattiva comprensione o applicazione delle istruzioni (anche se ricordiamo che è possibile costruire prove oggettive che prevedano più di una risposta), ma se il problema riguarda un item, è probabile che gli allievi che hanno segnato due crocette ritenessero le due risposte entrambe corrette, per cui l'analisi di quell'item dovrebbe considerare la possibilità di intervenire sull'item, anche in questo caso considerando l'entità dell'errore (con l'indice di facilità) e la sua distribuzione (indice di discriminatività).

5.6 I RISULTATI DEL GRUPPO DI ALLIEVI

Dopo aver analizzato il funzionamento dei quesiti che compongono la prova, passiamo alla verifica dei risultati ottenuti dal gruppo degli allievi. Si tratta di compilare la parte bassa della

scheda (Fig.5.9), in cui sono riportate alcune misure statistiche che descrivono i risultati della prova per l'intero gruppo. Illustreremo ora significato e modalità di ogni misura. Per esemplificare quanto detto faremo continuo riferimento alla Figura 5.8, in cui sono stati tabulati i risultati di un ipotetico gruppo di 19 studenti. Utilizzando un calcolatore e un foglio elettronico queste misure vengono calcolate velocemente (basta definire la serie di punteggi che si vogliono analizzare), ma per comprendere meglio il significato delle misure é utile ripercorrere tutta la procedura passo per passo, completando la tabella di Fig.5.8 con l'aiuto di una calcolatrice.

Il primo dato da ricavare é relativo al punteggio di **Moda**, cioè il punteggio ottenuto dagli studenti che ricorre più di frequente. Più che un calcolo si tratta di una osservazione, nella Fig.5.8 appare subito chiaro che il punteggio totale più frequente é il 5. Questo punteggio va riportato nel primo spazio delle statistiche descrittive, che si trova sul fondo del foglio di tabulazione delle risposte (nella figura che segue è rappresentata solo la parte bassa della tabella dati completa, visibile in Fig.5.8).

	1	3	4	4	4	5	5	5	5	5	6	6	6	7	7	8	8	9	10									

Statistiche descrittive

Moda_5_ Mediana___ Media____ Dev.St.____ C.d.v.______

Fig.5.9 - Indicazione del punteggio di Moda

Naturalmente se il gruppo fosse più ampio sarebbe più complessa l'osservazione del punteggio di moda, ma a questo proposito é utile ricordare che ogni foglio elettronico é in grado di estrarre facilmente il punteggio di moda da un elenco definito di punteggi. Può accadere che due o più punteggi risultino a pari merito per frequenza, ed in tal caso si parlerà di valori bimodali (o plurimodali).

Il punteggio di moda, velocemente osservabile, permette una prima stima dei risultati del gruppo, che dovrà comunque essere approfondita con le altre misure; un valore modale centrale (rispetto ai punteggi minimo e massimo) indica l'adeguatezza della prova per il gruppo di studenti, mentre un valore basso denota difficoltà di svolgimento e un valore alto facilità della prova; un valore modale forte (alta incidenza percentuale rispetto al totale dei punteggi) indica una certa omogeneità del gruppo, mentre la tendenza alla bimodalità, soprattutto di punteggi non vicini, denota squilibri fra gli allievi.

L'analisi delle frequenze dei punteggi oltre a indicare la moda permette di dare una visualizzazione grafica all'andamento dei punteggi. Si ricava un grafico a barre o a linea passante per tutte le frequenze dei punteggi, dal minimo teorico (0) al massimo teorico (il numero di quesiti della prova, nel nostro esempio 10), considerando anche eventuali punteggi a frequenza zero.

Il risultato é un grafico esemplificato nella Fig.5.10, che rappresenta la distribuzione dei punteggi della Fig.5.9. Come si può facilmente osservare il grafico raggiunge il punto più alto in corrispondenza della moda (5), mentre tende a scendere verso gli estremi.

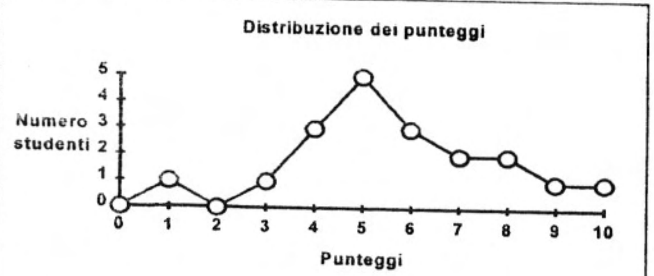

Fig.5.10-Grafico con la distribuzione dei punteggi

La seconda misura relativa alla statistica descrittiva è la **Mediana**, cioè il punteggio ottenuto dallo studente centrale nella distribuzione dei punteggi. Anche in questo caso l'osservazione diretta della tabella dei dati (Fig.5.8) permette di ricavare velocemente il valore della mediana: trattandosi di 19 studenti l'allievo centrale é il decimo, corrispondente al numero 1, che ha riportato il punteggio di 5. In questo caso la mediana coincide con la moda, ma non sempre questo avviene. Trascriviamo comunque anche questo dato nella parte bassa della tabella.

1	3	4	4	4	5	5	5	5	5	6	6	6	7	7	8	8	9	10

Statistiche descrittive

Moda_5_ Mediana_5_ Media____ Dev.St.____ C.d.v._____

Fig.5.11 - Indicazione del punteggio di Mediana

Una notazione tecnica: se il numero degli studenti é dispari, come nell'esempio, la mediana corrisponderà con il punteggio dello studente centrale; ma se il gruppo è composto da un numero di studenti pari, gli allievi centrali saranno due, per cui il valore della mediana corrisponderà alla media dei loro punteggi, cioè la loro somma divisa per due. Per esempio se il gruppo fosse stato composto da venti allievi avrei dovuto considerare i punteggi del decimo (5) e dell'undicesimo (nella nostra tabella 6), con un valore risultante della mediana di 5,5; a differenza della moda quindi la mediana può richiedere la cifra decimale.

L'importanza della mediana è data dal fatto che divide il gruppo in due metà, per cui è immediatamente osservabile quali studenti si trovino nella metà alta (o bassa) del gruppo. Inoltre se la prova è stata standardizzata, cioè ha delle statistiche descrittive di riferimento, ci dice immediatamente se il nostro gruppo ha ottenuto risultati migliori o peggiori del gruppo di riferimento della prova. Poiché il punteggio della mediana è quasi sempre costituito da un punteggio reale ottenuto da uno studente, e risente raramente della modifica di uno o pochi punteggi, viene considerato un dato robusto, utile soprattutto per i confronti, a differenza della media, che risulta più precisa (come vedremo), ma suscettibile di continue variazioni legate al variare dei punteggi.

La misura della mediana risulta utile soprattutto in rapporto al punteggio della media, che illustreremo tra poco, per cui rimandiamo l'analisi completa di questa misura. Integriamo però il grafico della distribuzione dei punteggi, tracciando una linea verticale continua in corrispondenza del valore della mediana; in questo modo abbiamo diviso la curva dei punteggi in due parti considerando l'uguale numerosità dei due gruppi.

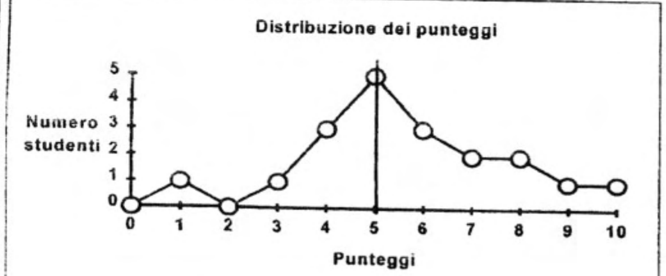

Fig.5.12-Grafico con la distribuzione dei punteggi+mediana

Per la terza misura di statistica descrittiva é necessario invece indicare una formula matematica, e quindi operare dei calcoli. Si tratta della **Media**, che corrisponde al punteggio medio ottenuto dal gruppo di allievi. La formula é comunque molto semplice:

$$\textbf{Media} = \frac{\text{Somma dei punteggi}}{\text{Numero degli allievi}}$$

Con l'aiuto di una calcolatrice si sommano i punteggi ottenuti da tutti gli allievi, e si divide il risultato per il loro numero. E' molto probabile che il risultato contenga cifre decimali, normalmente é sufficiente fermarsi al primo decimale, ma se la media é un numero basso é buona norma indicare anche i centesimi. Nel nostro esempio ecco il calcolo della media per i dati della tabella della solita Fig.5.8:

$$\text{Media Prova Esempio} = \frac{1+3+(4*2)+(5*5)+(6*3)+(7*2)+(8*2)+9+10}{19} = \frac{108}{19} = 5{,}68$$

Le parentesi si rendono necessarie quando un punteggio è stato raggiunto da più di uno studente, per cui conviene moltiplicare ogni punteggio per il numero di studenti che lo ha ottenuto e poi sommarlo agli altri punteggi (nell'esempio il punteggio 4 è stato raggiunto da due studenti). Il valore della media raggiunta dal gruppo di allievi nella prova esempio é dunque di 5,68, da trascrivere nello spazio relativo delle statistiche descrittive.

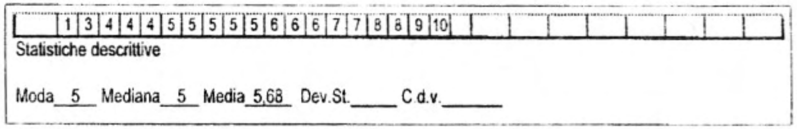

Fig.5.13 - Indicazione del punteggio di Media

Il semplice dato ci può dare solo alcune indicazioni: si può confrontare con la media teorica, cioè il valore centrale dei punteggi possibili, nel nostro caso 5,5, perché bisogna considerare anche il punteggio zero, cioè l'allievo che non risponde ad alcuna domanda, e poiché la media del gruppo é maggiore, si può stimare una tendenziale facilità della prova per il gruppo di studenti.

Più utile risulta il confronto fra la media e la mediana: per renderlo più evidente completiamo ancora il grafico tracciato in precedenza con un'altra linea verticale, di diverso colore o tratto (nell'esempio è tratteggiata), corrispondente al valore della media.

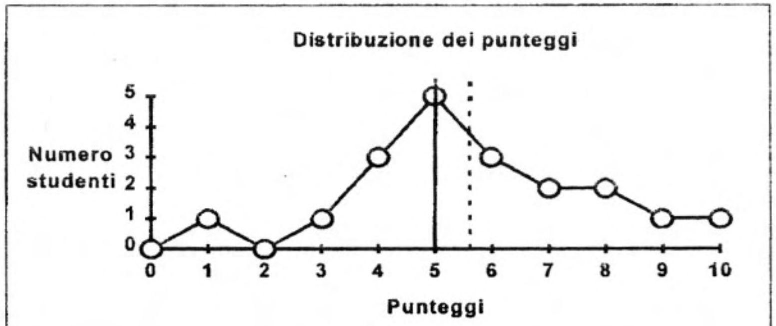

Fig.5.14-Grafico con la distribuzione dei punteggi+mediana e media

Anche questa seconda linea dividerà la curva di frequenze in due parti, ma non considerando la numerosità degli studenti, bensì i loro punteggi. Le due linee verticali, corrispondenti a mediana e media, si troveranno in una posizione reciproca, e si potranno verificare tre tipi di posizione, visualizzabili nella Fig.5.15 (le frequenze dei punteggi sono inventate).

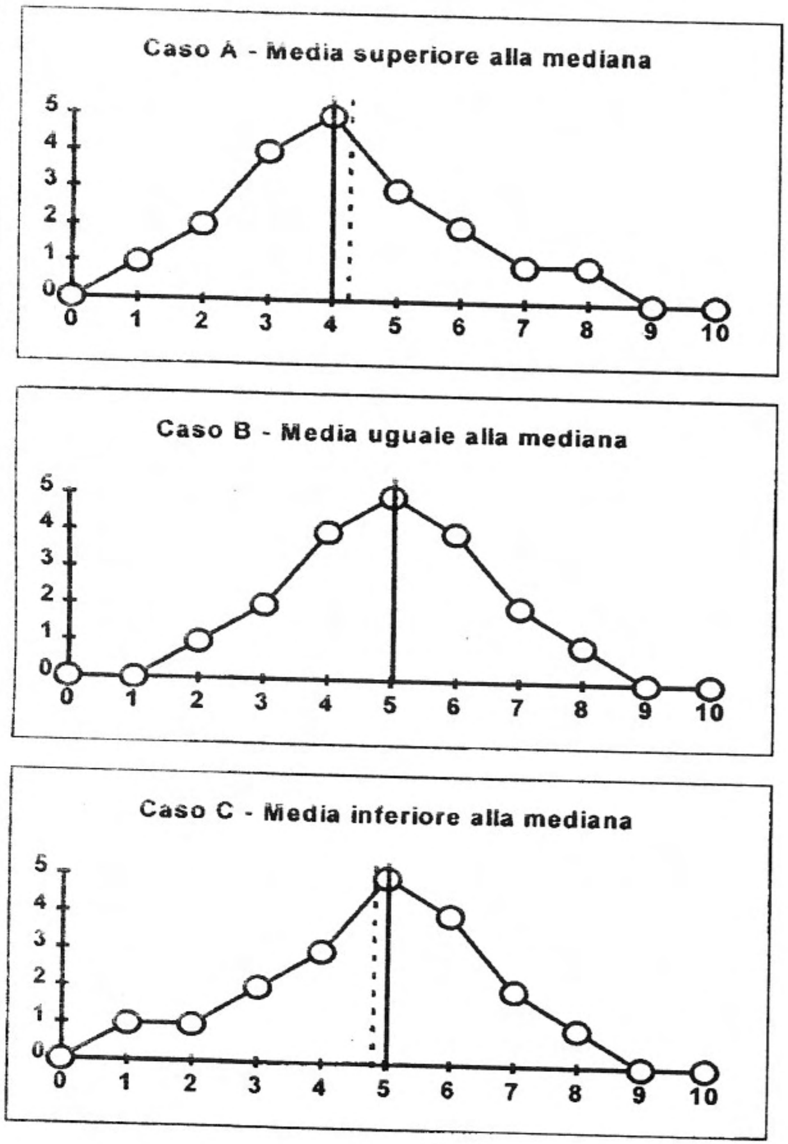

Fig.5.15 - Posizione reciproca mediana e media - Casi possibili

Nel primo caso (A) la media é più alta della mediana, cioè si trova alla sua destra; la curva delle frequenze dei punteggi risulta necessariamente asimmetrica, con una maggiore densità di punteggi bassi, per cui si dice ad asimmetria positiva, cioè la prova è risultata difficile per la maggior parte del gruppo che l'ha svolta. E' questo il caso a cui più si avvicina il nostro esempio (vedi Fig.5.14)

Nel caso centrale (B) la media e la mediana coincidono, o tendono a coincidere: la distribuzione dei punteggi tende ad essere simmetrica, e corrisponde alla cosiddetta distribuzione normale, rappresentata da una curva con forma a campana; si tratta di una situazione molto frequente per grandi rilevazioni e gruppi medi, con una maggiore densità per i punteggi centrali e una costante diminuzione degli estremi, sia positivo che negativo.

Infine il terzo caso (C), opposto al primo, in cui la linea della media si trova a destra della mediana; in tal caso parliamo di asimmetria negativa e di relativa facilità della prova per il gruppo di riferimento.

5.7 L'OMOGENEITÀ DEL GRUPPO

Finora abbiamo considerato i risultati ottenuti dal gruppo come se fosse un singolo organismo, e le statistiche descrittive servono proprio a conoscere l'andamento complessivo del gruppo, ma non bisogna dimenticare che un gruppo é formato da un insieme di individui, e che due gruppi con risultati medi simili potrebbero in realtà essere molto diversi nella loro composizione, cioè nei risultati individuali ottenuti dai singoli allievi.

Per fare un esempio molto semplice, due gruppi di persone potrebbero entrambi essere mediamente alti 170 centimetri, ma mentre il primo presenta tutti individui vicino alla media, diciamo con un limite minimo di 165cm. (il componente più basso) e massimo di 175cm. (il più alto), il secondo gruppo é composto da individui molto bassi, anche 150cm., e molto alti, fino a 190cm. La media é uguale o simile per i due gruppi, ma é ottenuta con misure molto diverse. Il primo gruppo risulta tendenzialmente omogeneo (misure simili e vicine), mentre il secondo é caratterizzato da una accentuata eterogeneità. La misura che ci indica il livello di omogeneità del gruppo é la **deviazione standard**, che si basa sul principio di considerare, facendone la somma, gli scarti dei singoli punteggi ottenuti dagli studenti rispetto alla media. La formula per il calcolo delle deviazione standard può apparire più complessa, ma si tratta solo di un'impressione:

$$\textbf{Deviazione standard} = \sqrt{\frac{\Sigma(Xi - X)^2}{n}}$$

Il simbolo Σ indica l'operazione di sommatoria, cioè la somma di una serie di operazioni, nel nostro caso le sottrazioni fra ogni singolo punteggio ottenuto dagli studenti (Xi) e il punteggio di media (X), con il risultato di ogni singola sottrazione elevata al quadrato. La somma di tutte queste sottrazioni viene divisa per il numero (n) degli allievi che compongono il gruppo. Infine si estrae la radice quadrata del risultato. Quest'ultima operazione, come i precedenti quadrati delle sottrazioni, sono un semplice artificio matematico per ovviare a un problema di calcolo, che apparirà evidente nell'esempio di deviazione standard applicato ai risultati del nostro gruppo ipotetico. Consideriamo in particolare i punteggi totali di ogni studente, riportati nell'ultima riga della tabella dati (vedi Fig. 5.13).

Utilizziamo questi punteggi per effettuare le 19 sottrazioni previste dalla formula per il calcolo della deviazione standard, cioé sottraiamo da ogni punteggio la media, che ricordiamo é di 5,68:

1-5,68 = -4,68
3-5,68 = -2,68
4-5,68 = -1,68
5-5,68 = -0,68
6-5,68 = 0,32
7-5,68 = 1,32
8-5,68 = 2,32
9-5,68 = 3,32
10-5,68 = 4,32

E' abbastanza semplice constatare che per tutti i punteggi inferiori alla media il risultato della sottrazione sarà negativo; se a questo punto si effettuasse la sommatoria (somma dei risultati delle sottrazioni) il totale sarebbe 0, perché è stato utilizzato come riferimento il punteggio di media, e il calcolo della deviazione standard si fermerebbe qui. Ecco perché é necessario elevare al quadrato i risultati delle sottrazioni; i quadrati sono tutti numeri

positivi, quindi la loro somma sarà sicuramente un numero positivo diverso da 0. Per i punteggi ottenuti da più di uno studente è possibile, dopo il calcolo, moltiplicare il risultato per il numero di studenti che ha ottenuto quel punteggio (è il caso, nel nostro esempio, dei punteggi compresi fra 4 e 8).

Punteggio grezzo-media	Risultato da elevare al quadrato	Risultato	Numero studenti	Risultato per la sommatoria
$(1-5{,}68)^2$	$-4{,}68^2$	21,90	1	21,90
$(3-5{,}68)^2$	$-2{,}68^2$	7,18	1	7,18
$(4-5{,}68)^2$	$-1{,}68^2$	2,82	3	8,46
$(5-5{,}68)^2$	$-0{,}68^2$	0,46	5	2,3
$(6-5{,}68)^2$	$0{,}32^2$	0,10	3	0,30
$(7-5{,}68)^2$	$1{,}32^2$	1,74	2	3,48
$(8-5{,}68)^2$	$2{,}32^2$	5,38	2	10,76
$(9-5{,}68)^2$	$3{,}32^2$	11,02	1	11,02
$(10-5{,}68)^2$	$4{,}32^2$	18,66	1	18,66
			Totale	**84,06**

Fig.5.16 - Tabella di calcolo per la sommatoria necessaria per la Deviazione standard

A questo punto posso effettuare la sommatoria dei risultati delle sottrazioni (al quadrato), per cui la formula di calcolo della deviazione standard, applicata all'esempio, risulta:

$$\text{Deviazione standard gruppo} = \sqrt{\frac{84{,}06}{19}} = \sqrt{4{,}43} = 2{,}10$$

La deviazione standard della classe, per quanto riguarda la prova, é quindi di 2,10, e possiamo continuare a riempire la nostra tabella dati nella parte bassa.

	1	3	4	4	4	5	5	5	5	5	6	6	6	7	7	8	8	9	10								

Statistiche descrittive

Moda 5 Mediana 5 Media 5,68 Dev.St. 2,10 C.d.v.

Fig.5.17- Statistiche descrittive con Deviazione standard

Abbiamo detto che questa misura si riferisce al livello di omogeneità del gruppo. Ma a quale livello comincia l'omogeneità? Per rispondere dobbiamo calcolare l'indice successivo alla deviazione standard, cioè il **coefficiente di variazione** (C.d.v.).

Si tratta di un'operazione molto semplice:

$$\textbf{Coefficiente di variazione} = \frac{\text{Deviazione standard}}{\text{Media}} * 100$$

La deviazione standard è una misura relativa alla media, e il coefficiente di variazione é in pratica la percentuale della deviazione standard rispetto al punteggio di media. Quando il coefficiente è inferiore a 10 si può parlare di un gruppo estremamente omogeneo, mentre un coefficiente superiore a 20 indica una forte eterogeneità del gruppo. E' il caso del nostro esempio, infatti se calcoliamo il coefficiente di variazione con i dati del nostro gruppo otteniamo:

$$\text{Coefficiente di variazione gruppo} = \frac{2{,}10}{5{,}68} * 100 = 37$$

Possiamo così completare con il valore ottenuto la nostra tabella dati

1	3	4	4	4	5	5	5	5	5	6	6	6	7	7	8	9	10

Statistiche descrittive

Moda 5 Mediana 5 Media 5,68 Dev.St. 2,10 C.d.v. 37

Fig.5.18 - Statistiche descrittive con Coefficiente di variazione

Il livello di eterogeneità è altissimo, e d'altra parte si poteva intuire dalla distribuzione dei punteggi, con diversi allievi vicini al massimo (uno addirittura a 10), mentre alcuni compagni hanno punteggi molto bassi. Nella realtà é difficile incontrare gruppi così poco omogenei, nella maggior parte dei casi il coefficiente di variazione si trova fra 10 e 20. Sicuramente una tendenza all'omogeneità facilità il lavoro di formazione per il docente, poiché i livelli di abilità degli allievi risultano vicini. Al contrario l'ipotetico formatore del nostro gruppo esempio dovrebbe fare i conti con livelli di abilità estremamente differenziati, con conseguenti complicazioni per un'attività didattica che risulti efficace per tutti.

Durante l'attività di formazione risulta molto utile il confronto diacronico dei coefficienti di variazione ottenuti dal gruppo in diverse prove riguardanti abilità simili: se tendono a diminuire vuol dire che il gruppo sta diventando sempre più omogeneo per quell'area di abilità, con conseguenti vantaggi per la didattica; al contrario se cresce l'eterogeneità vuol dire che le differenze originarie del gruppo si stanno ampliando, la distanza fra bravi e meno bravi cresce, il gruppo tende a spaccarsi.

Anche il confronto sincronico, effettuato fra prove di ambiti diversi realizzate in tempi ravvicinati, può fornire interessanti indicazioni; se l'omogeneità del gruppo varia sensibilmente in diversi tipi di prove l'intervento didattico potrà valorizzare gli elementi di omogeneità per accrescere il senso del gruppo nella classe, mentre per gli aspetti che risultano meno omogenei sarà possibile intervenire con maggiore consapevolezza delle differenze individuali.

Da notare che l'omogeneità del gruppo non dipende dalla qualità dei risultati: ci possono essere classi con medie alte, ma forti differenze interne, o classi scadenti molto compatte, per cui potrebbe risultare più semplice elaborare un itinerario di formazione proprio per la classe scadente, i cui studenti però si trovano tutti sullo stesso piano.

Per tornare a una rappresentazione grafica dell'obiettivo auspicabile di ogni intervento didattico, la curva generata dalla distribuzione normale dei punteggi (vedi Fig.5.15 - Caso B), che caratterizza un gruppo medio prima dell'intervento, dovrebbe trasformarsi e diventare una curva a forte asimmetria negativa (vedi Fig.5.19), che denota abbondanza di punteggi alti e omogeneità del gruppo.

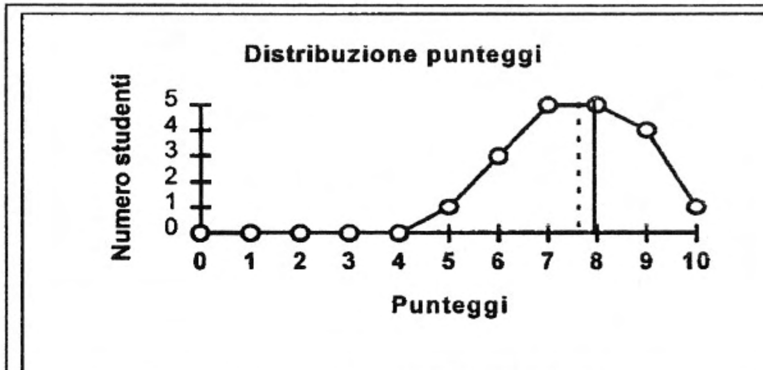

Fig.5.19-Distribuzione dei punteggi obiettivo dopo un intervento didattico

Una volta ottenuta la distribuzione dei punteggi per una prova di tipo oggettivo si deve risolvere il problema di trasformare i punteggi in giudizi. Proseguendo nel nostro esempio, quale punteggio un allievo deve ottenere per essere considerato sufficiente? E con quale punteggio possiamo dire che il risultato è ottimo? O ancora, al di sotto di quale punteggio la prova deve essere considerata del tutto insoddisfacente? Il buon senso può fornire delle risposte a questi interrogativi, ma ci possono essere diversi tipi di buon senso. Si può decidere che la sufficienza scatti a 5 punti, il giusto mezzo dei punteggi possibili; oppure a 6, similmente alla sufficienza tradizionale scolastica (la prova utilizzata come esempio prevede proprio 10 punti, come i voti); oppure a 7 (o 4) perché in definitiva la prova è abbastanza facile (o difficile). Come si vede ogni soglia può trovare una definizione di buon senso, per cui gli stessi punteggi di una stessa prova possono avere per insegnanti diversi valori diversi (come sono diversi i buoni sensi).

Anche in questo caso i numeri ci possono aiutare; l'esperienza docimologica ha permesso di definire la distribuzione di punteggi basata su una scala pentenaria, composta cioè da cinque giudizi normalmente definiti con le prime cinque lettere dell'alfabeto, per cui si va da A (punteggi fortemente positivi) a E (punteggi del tutto negativi). Il problema è quello della definizione delle soglie, cioè i punti di passaggio da una lettera all'altra.

Il problema viene risolto considerando i singoli risultati nel contesto dell'intero gruppo, cioè ogni punteggio risulterà negativo, sufficiente o positivo rispetto ai risultati dell'intero gruppo. Come è facilmente intuibile questa procedura introduce una serie di problemi di "giustizia" didattico-formativa, poiché le valutazioni ottenute con la distribuzione dei punteggi risultano sempre relative al gruppo esaminato. Rimandiamo per ora ogni puntualizzazione su questo aspetto, pur rimanendone consapevoli, per approfondire le modalità tecniche di attuazione della distribuzione pentenaria; anticipiamo solo che i criteri per stabilire soglie di sufficienza ed eccellenza sono sempre relativi, il vero problema è la definizione del campo di relazione e le modalità di calcolo.

Per la distribuzione pentenaria dei punteggi ritorniamo ai risultati della nostra classe esempio (vedi Fig.5.17).

Il punto cardine per definire il valore dei punteggi è costituito dal punteggio medio, nel nostro caso 5,68. Poiché una scala pentenaria è costituita da 5 giudizi, possiamo dire che il punteggio medio rientra nel giudizio centrale, cioè C. Ma nella realtà nessun allievo ha ottenuto il punteggio medio, alcuni hanno 5 punti, altri 6: anche loro possono ricevere il giudizio C o passano alla fascia successiva o precedente? Per definire il problema della soglia dobbiamo utilizzare la misura che definisce l'omogeneità del gruppo, cioè la Deviazione standard. Possiamo dire che il giudizio C corrisponde a una fascia di ampiezza uguale alla deviazione standard, nel nostro caso 2,10. Poiché il centro della fascia C è definito dal punteggio medio, i limiti estremi di questa fascia si possono calcolare togliendo (e aggiungendo) dalla media il valore di mezza deviazione standard, cioè 1,05. Compiendo le due operazioni si ottengono i valori di 4,63 (limite inferiore della fascia C) e 6,73 (limite superiore). Pertanto i punteggi superiori a 4,63 e inferiori a 6,73 rientrano nella fascia C e possono essere considerati sufficienti. Nel nostro caso si tratta degli otto allievi che hanno ottenuto 5 o 6.

Partendo dalle soglie così calcolate proseguiamo la definizione delle altre fasce. La fascia B avrà come limite inferiore la soglia alta della fascia C, cioè 6,73, mentre per il limite superiore bisognerà aggiungere un'altra deviazione standard, quindi 6,73+2,10=8,83. In questa fascia, che può essere considerata quella di un buon risultato, rientrano i quattro allievi che hanno ottenuto i punteggi 7 e 8. Per la fascia A, risultato ottimo, non serve calcolare il limite superiore, evidentemente costituito dal punteggio massimo della prova. Nel nostro caso in questa fascia rientrano due studenti che hanno ottenuto 9 e 10 punti.

All'opposto definiamo i limiti della fascia D, cioè i risultati insufficienti: il limite superiore corrisponde alla soglia inferiore della fascia C, cioè 4,63, mentre per il limite inferiore dobbiamo scendere di una deviazione standard, quindi 4,63-2,10=2,53. Rientrano in questa fascia i quattro studenti con punteggi 3 e 4. Infine la fascia E, per i punteggi inferiori a 2,53 (fino al minimo teorico di 0), con un solo studente che ha il punteggio 1.

Per riepilogare componiamo il seguente grafico a barre, in cui ogni colonna è relativa a una fascia di livello.

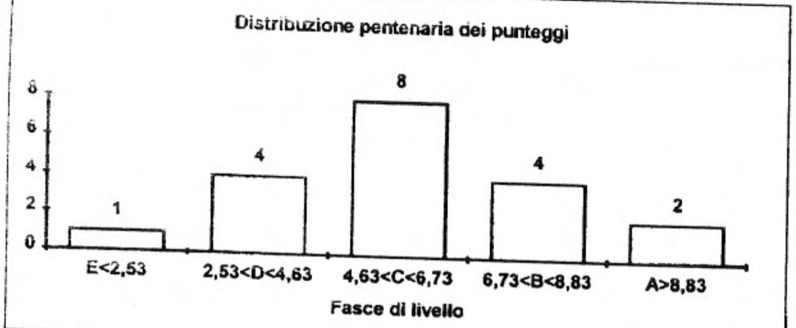

Fig.5.20 - Distribuzione pentenaria dei punteggi

Si può facilmente notare che la curva di distribuzione pentenaria assume una caratteristica forma a campana, molto simile a quella già vista nella distribuzione di frequenza dei punteggi (vedi Fig.5.15 - Caso B). In genere la fascia centrale è quella più rappresentata, mente le due fasce estreme risultano ridotte e tendono all'equivalenza fra di loro. La statistica ci conferma l'esistenza di una distribuzione normale dei punteggi fra le cinque fasce di livello, ovviamente simmetrica verso gli estremi. Nel grafico che segue sono rappresentate le percentuali normali di ogni fascia, cioè i punteggi che possiamo attenderci in un gruppo medio.

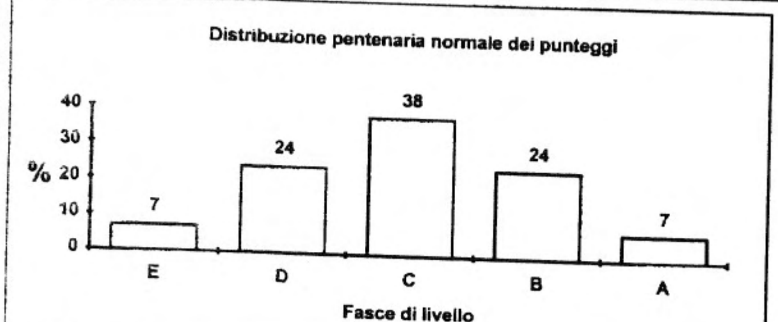

Fig.5.21 - Distribuzione pentenaria dei punteggi secondo una curva normale

Come si può notare l'andamento del nostro gruppo esempio è molto simile ad una distribuzione normale, ecco infatti il confronto fra le percentuali di ogni fascia.

	Fascia E	Fascia D	Fascia C	Fascia B	Fascia A
Gruppo esempio	5,26	21,05	42,1	21,05	10,54
Distribuzione normale	7	24	38	24	7

Fig.5.22 - Confronto punteggi pentenari fra gruppo esempio e distribuzione normale

Come abbiamo già notato nel paragrafo precedente, una distribuzione normale, anche dei punteggi pentenari, è la più frequente in un gruppo prima di un intervento didattico; l'obiettivo di un intervento formativo è un diffuso miglioramento dei risultati, ma in particolare una forte contrazione dei punteggi relativi alle fasce insufficienti. Se infatti applichiamo la distribuzione pentanaria ai dati relativi al grafico della Figura 5.18 (punteggi ottenuti dopo un ipotetico intervento didattico) otteniamo la seguente distribuzione:

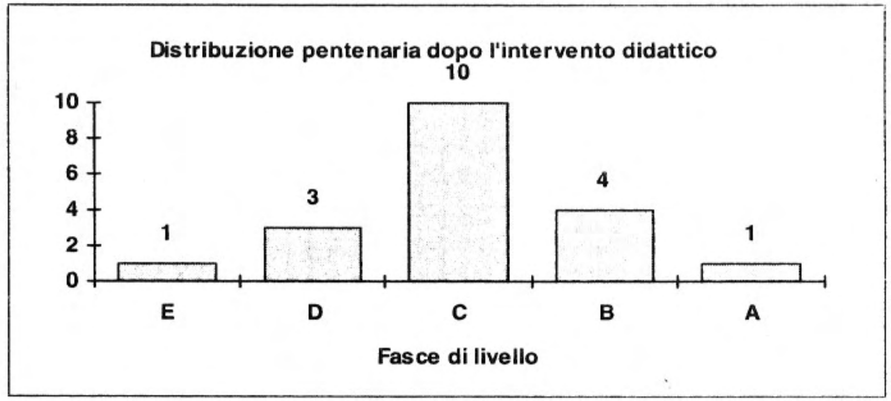

Fig. 5.23 - Distribuzione pentenaria possibile dopo un intervento didattico positivo

Oltre ad un miglioramento assoluto dei risultati (la media è passata da 5,68 a 7,57) è cambiato l'equilibrio fra punteggi insufficienti (in diminuzione) e sufficienti (aumentati), con un netto addensamento (più della metà degli allievi) nella fascia centrale. Il dato può sembrare penalizzante, paradossalmente, per gli studenti più bravi, che non aumentano, ma bisogna considerare che è aumentata, insieme alla media, anche l'omogeneità del gruppo, cioè è diminuita la deviazione standard (in questo esempio è passata da

2,10 a 1,27) per cui i livelli di abilità degli allievi risultano ravvicinati, con ovvi vantaggi per le fasce intermedie C e B (che si avvicinano all'eccellenza), ma anche per la fascia D, la cui abilità non è poi così lontana dalla sufficienza.

5.9 STANDARDIZZAZIONE DEI PUNTEGGI

Abbiamo visto come sia possibile, in base alla distribuzione dei punteggi ottenuti dagli allievi, stabilire le soglie di sufficienza ed eccellenza per i punteggi ottenuti in una prova, e di conseguenza come sia possibile trasformare il punteggio in un giudizio di valutazione sintetico (le fasce della distribuzione pentenaria). Il limite più importante di questa procedura è lo stretto legame esistente fra il punteggio e la prova stessa, sia per come è costruita che per i risultati del gruppo che l'ha svolta. Questo rende difficilmente confrontabili punteggi ottenuti in prove diverse: se le prove sono costituite da un numero diverso di quesiti è evidente che il confronto dei punteggi grezzi non può essere considerato valido; se anche i punteggi confrontati si riferiscono a prove con la stessa struttura, la distribuzione dei punteggi può essere estremamente diversa, in particolare medie e deviazioni standard possono differire di molto.

Per ovviare a questo problema, e permettere quindi il confronto fra punteggi ottenuti in prove diverse, si può procedere ad una standardizzazione dei punteggi, cioè collegarli a riferimenti stabili. Per riconoscere immediatamente un punteggio standard da un punteggio grezzo viene usata la notazione *z*, per cui 1,2*z* si riferisce a un punteggio standard in punti *z*.

La trasformazione di un punteggio grezzo in punti *z* risulta abbastanza semplice dal punto di vista matematico:

$$\textbf{punteggio } z = \frac{\text{punteggio grezzo - punteggio medio}}{\text{deviazione standard}}$$

In pratica da ogni punteggio grezzo si sottrae la media registrata per la prova e si divide il risultato per la deviazione standard relativa. Applichiamo la trasformazione ai punteggi della prova esempio già incontrata nella procedura dell'item analisi (ricordiamo che la media era 5,68 e la deviazione standard 2,10; sono stati raggruppati gli allievi che hanno ottenuto lo stesso punteggio).

Studenti	punteggio grezzo	operazione	punti *z*
7	1	(1-5,68)/2,10	-2,23
5	3	(3-5,68)/2,10	-1,28
12-9-17	4	(4-5,68)/2,10	-0,8
6-19-11-3-1	5	(5-5,68)/2,10	-0,32
15-18-4	6	(6-5,68)/2,10	0,15
13-2	7	(7-5,68)/2,10	0,63
10-14	8	(8-5,68)/2,10	1,10
8	9	(9-5,68)/2,10	1,58
16	10	(10-5,68)/2,10	2,06

Fig.5.24 - Tabella di trasformazione dei punteggi grezzi in punti *z*

Come si può notare i punti *z* si riferiscono alla media (lo 0 coincide proprio con il valore medio) e il loro incremento è determinato dalla deviazione standard (un punto *z* corrisponde al valore della deviazione standard, 2*z* a due deviazioni standard e così di seguito), che rappresenta quindi l'unità di misura dei punti *z*. Di conseguenza i punti *z* variano fra -3 e +3 (in teoria sono possibili anche punteggi superiori e inferiori, ma con probabilità vicine allo 0), a prescindere dal numero di quesiti che compongono la prova. La lettura dei punti *z* permette anche di attribuire ad ogni allievo la fascia pentenaria di appartenenza: poiché ogni fascia è larga esattamente una deviazione standard, e ricordando che la fascia C è centrata sulla media, tutti gli allievi compresi fra -0,50*z* e 0,50*z* ottengono C, quelli compresi fra 0,50*z* e 1,50*z* ottengono B e così di seguito. Ecco la trasformazione applicata al nostro esempio.

Studente	punteggio grezzo	punti *z*	fascia pentenaria
7	1	-2,23	E
5	3	-1,28	D
12-9-17	4	-0,8	D
6-19-11-3-1	5	-0,32	C
15-18-4	6	0,15	C
13-2	7	0,63	B
10-14	8	1,10	B
8	9	1,58	A
16	10	2,06	A

Fig. 5.25 - Tabella di trasformazione dei punti *z* in giudizio pentenario

La standardizzazione dei punteggi può quindi sostituire anche la procedura di distribuzione pentenaria, fornendo immediatamente il livello di rendimento di ogni allievo.

Poiché i punteggi standardizzati fanno riferimento a valori relativi al gruppo è evidente che il confronto fra prove diverse può avvenire solo se coincide il gruppo che le ha svolte: per un allievo ogni variazione di punteggio z significa in pratica una diversa collocazione di rendimento all'interno del gruppo di cui fa parte, cioè, in genere, la sua classe.

Un aspètto poco simpatico dei punti standardizzati z è che prevedono valori negativi o anche il valore 0, riferito peraltro a un punteggio perfettamente in media, è quindi positivo.

Volendo ovviare a questo inconveniente (basta pensare alla comunicazione con i colleghi o con gli allievi stessi) è possibile applicare un facile artificio matematico per ottenere punteggi standardizzati sempre espressi con valori positivi. Parleremo in questo caso di punteggio standardizzato T, così calcolabile:

punteggio T = $50 + 10z$

In pratica lo 0 dei punti z viene spostato a 50: tutti i valori inferiori a 50T corrispondono ai punti z negativi. Ricordando i limiti di variazione dei punti z è evidente che i punti T non potranno mai essere negativi. Applichiamo la trasformazione al nostro esempio.

Studente	punt. grezzo	punti z	operazione	punti T	fascia
7	1	-2,23	50+10(-2,23)	27,7	E
5	3	-1,28	50+10(-1,28)	37,2	D
12-9-17	4	-0,8	50+10(-0,8)	42	D
6-19-11-3-1	5	-0,32	50+10(-0,32)	46,8	C
15-18-4	6	0,15	50+10(0,15)	51,5	C
13-2	7	0,63	50+10(0,63)	56,3	B
10-14	8	1,10	50+10(1,10)	61	B
8	9	1,58	50+10(1,58)	65,8	A
16	10	2,06	50+10(2,06)	70,6	A

Fig.5.26 - Tabella di trasformazione dei punti z in punti T

Come per i punti z anche per i punti T è semplice definire i confini delle fasce pentenarie: fra 45T e 55T fascia C, fra 55T e 65T fascia B e via di seguito.

Il trattamento dei punteggi con un programma di foglio elettronico permette una rapida trasformazione in punti standardizzati (si possono ottenere i punti T con un solo passaggio matematico unendo le formule sopra indicate) ed un rapido confronto dei rendimenti di un allievo anche in una lunga serie di prove diverse, accertando il suo miglioramento (o peggioramento) all'interno del gruppo classe.

5.10 QUELLO CHE LA STATISTICA NON PUÒ DIRE

Con queste due unità abbiamo cercato di presentare alcune procedure utili per valutare i materiali costruiti per ottenere misurazioni di tipo oggettivo (item analisi) e per valutare gli allievi che svolgono le stesse prove (statistiche descrittive). La distinzione sembra far riferimento a una categorizzazione generale della statistica, suddivisa fra inferenziale e descrittiva. In realtà i due aspetti non sono divisi in modo così netto: la procedura di item analisi ha come scopo principale la comprensione (inferenziale) della funzionalità di un mezzo, indicando dove c'è qualcosa che non va ed è il caso di intervenire, modificando o sopprimendo quesiti; ma l'insieme dei dati sicuramente descrive la funzionalità della prova (che può essere confrontata con prove simili) e in ultima analisi descrive l'abilità del formatore nella costruzione o nella scelta dello strumento in relazione al suo gruppo di allievi. D'altra parte anche le statistiche descrittive, la cui ovvia funzione è ben chiara nel nome, permettono di inferire elementi circa l'andamento del gruppo di allievi, soprattutto per la possibilità di confronti nel tempo o con gruppi simili.

Le due operazioni (descrizione e inferenza) devono comunque essere guidate dalla sensibilità didattica del formatore, che deve essere in grado di leggere adeguatamente le informazioni fornite dall'indagine statistica. Abbiamo accennato in diverse occasioni alle diverse spiegazioni possibili per un indice di facilità alto o una deviazione standard contenuta; è il formatore che elabora i dati che deve scegliere l'ipotesi di spiegazione più probabile, incrociando i numeri con tutte le altre possibili osservazioni della realtà didattica che è riuscito a raccogliere, nonché confrontando i dati con le ipotesi di partenza, che dovrebbero aver guidato la programmazione del curricolo, la realizzazione delle attività e la scelta degli strumenti di verifica.

La statistica può fornire numeri, indici, rapporti, con indicazioni relative a esperienze consolidate, ma non può spiegare da sola i fenomeni; giornalisticamente parlando può indicare il che cosa, ma non spiegare il perché. Questo è competenza del formatore, che deve avere la capacità analitica di raccogliere informazioni riguardanti i suoi allievi e la capacità sintetica di trasformare la raccolta dati in decisione formativa.

Le indicazioni fornite in questo capitolo rappresentano solo parzialmente il contributo che l'analisi statistica può fornire alla formazione, ma le procedure presentate sono sicuramente tra le più utili ed utilizzabili in campo formativo. Per eventuali approfondimenti rimandiamo ai titoli segnalati nella bibliografia, ricordando che un'altra abilità importante per un tecnico di valutazione è la lettura e l'interpretazione dei dati statistici riguardanti l'istruzione: la conoscenza delle tecniche di campionamento, delle modalità di raccolta e presentazione dei dati (ad esempio le statistiche periodiche sull'istruzione dell'ISTAT), la lettura critica delle interpretazioni fornite sui dati non possono essere trattate in questa sede, ma dovrebbero far parte a pieno titolo delle competenze di un formatore, per riflettere in modo attivo (utile cioè per la propria esperienza professionale) sulla crescente massa di dati che la pubblicistica mette a disposizione degli operatori.

Nel paragrafo 5.8 abbiamo lasciata sospesa una questione riguardante la relativizzazione delle valutazioni, ed in particolare la necessità di riferire i punteggi grezzi all'andamento del gruppo di allievi. Una obiezione abbastanza scontata è che il punteggio grezzo di uno studente potrebbe portare a valutazioni diverse se inserito in un gruppo di livello medio-alto o basso (nel paragrafo 5.8 parlavamo appunto di "giustizia scolastica"). Le risposte possibili a questa obiezione sono diverse:

- se la prova è in ingresso, l'allievo (e il formatore) ha tutto da guadagnare nel conoscere la sua posizione relativa all'interno del gruppo, in tal modo l'intervento didattico potrà adeguarsi alle sue capacità e a quelle degli altri allievi;
- se la prova è formativa (*in itinere*), deve essere costruita tenendo conto del percorso didattico effettivamente svolto, per cui il livello medio del gruppo dovrebbe aver influenzato l'allievo prima della prova; cioè nel gruppo più abile le spiegazioni, i materiali, i lavori di gruppo saranno stati di livello più alto ed avranno influenzato in tal senso la partecipazione, e di conseguenza il rendimento, del singolo allievo;
- se la prova è sommativa (finale), il rischio di ingiustizia è più concreto (se non altro perché l'esito potrebbe essere una promozione, un diploma, una qualifica), ma questa consapevolezza permette di affrontare con chiarezza il problema: le prove finali dovrebbero essere, per quanto possibile, di tipo standardizzato, cioè riferito a un gruppo più ampio di allievi.

Per quest'ultimo aspetto ecco alcune possibili soluzioni, che hanno lo scopo di aumentare il valore delle indicazioni provenienti dall'analisi statistica delle prove;

- se due o più gruppi di allievi svolgono contemporaneamente la stessa prova è possibile considerarli un solo gruppo (e una sola serie di statistiche descrittive), superando il criterio delle valutazioni relative a una sola classe; naturalmente si perdono le informazioni relative all'omogeneità del gruppo (però si possono elaborare i dati anche suddivisi per gruppo) e l'aggregazione resta possibile solo per le prove che misurino abilità trattate in modo equivalente nei due o più gruppi considerati;
- l'aggregazione dei dati riguardanti più gruppi può essere anche diacronica, quando la stessa prova è somministrata a gruppi diversi, ma in anni successivi; restano validi gli accorgimenti indicati nel precedente punto, con l'aggiunta che la riproposizione costante della stessa prova permette di raccogliere utili informazioni circa l'evoluzione delle abilità degli allievi e dei formatori: è ragionevole ipotizzare che con il passare degli anni una prova diventi sempre più facile;
- si possono utilizzare prove standardizzate su una popolazione ampia, di estensione regionale o nazionale, corredate da statistiche descrittive di riferimento con cui confrontare i risultati del proprio gruppo di allievi; è questa sicuramente un'ottima soluzione per comprendere i progressi del gruppo non solo in relazione al proprio curricolo, ma rispetto a standard di riferimento più ampi, naturalmente se la prova è compatibile con il curricolo effettivamente svolto.

L'ultima soluzione indicata è purtroppo la meno praticabile, perché non sono in genere disponibili prove standardizzate su popolazioni di una certa ampiezza. La tendenza in atto nel sistema educativo italiano è sicuramente questa, come dimostra la recente definizione di un Servizio Nazionale per la Qualità dell'Istruzione presso il CEDE, che ha avviato l'archiviazione e la validazione di prove oggettive per tutti i livelli scolastici. Nel frattempo chi si occupa di istruzione e formazione in Italia deve accontentarsi nella maggior parte dei casi di autocostruirsi prove standardizzate, possibilmente in collaborazione con colleghi dello stesso corso o di discipline affini.

E' questo sicuramente uno dei compiti più importanti di un tecnico di valutazione degli apprendimenti, e il tentativo di questo capitolo, e più in generale di tutto il volume, è stato proprio quello di fornire gli strumenti per controllare in modo consapevole tutte le procedure che presiedono alla definizione di una valutazione corretta e utile per una formazione professionale efficace.

BIBLIOGRAFIA

Asquini G., *Il foglio di calcolo per l'analisi delle prove a risposta chiusa*, in Cadmo, n.13-14, 1997.

Benvenuto G. - Lastrucci E. - Salerni A., *Leggere per capire*, Roma, Anicia, 1995.

Coggi C. - Calonghi L., *Elementi di statistica per la ricerca scolastica*, Teramo, Lisciani e Giunti, 1992.

Domenici G., *Le prove strutturate di conoscenza*, Teramo, Lisciani e Giunti, 1992.

Domenici G., *Manuale della valutazione scolastica*, Bari, Laterza, 1993.

Gattullo M. - Giovannini L., *Misurare e valutare l'apprendimento nella scuola media*, Milano, Bruno Mondadori, 1989.

Lombardo E., *I dati statistici in pedagogia. Esplorazione ed analisi*, Firenze, La Nuova Italia,1993.

Lucisano P. (a cura di), *Lettura e comprensione*, Torino, Loescher, 1989.

Vertecchi B., *Manuale della valutazione. Analisi degli apprendimenti*, Roma, Editori Riuniti, 1984.

Vertecchi B., *Decisione didattica e valutazione*, Firenze, La Nuova Italia, 1993.