

La validazione delle prove dei gruppi disciplinari (TECO-D)

Sommario

INTRODUZIONE	2
CARATTERISTICHE DELLA MATRICE DATI PER L'ITEM ANALYSIS	2
ITEM ANALYSIS E FEEDBACK ANVUR.....	3

INTRODUZIONE

A conclusione della somministrazione pilota della prova predisposta da ogni Gruppo di Lavoro (GdL), l'ANVUR redige una relazione tecnica con l'obiettivo di fornire riscontri puntuali sulla validità e affidabilità sia della prova nel suo complesso sia del funzionamento dei singoli item.

La relazione include:

- una breve spiegazione delle procedure di analisi;
- gli indicatori principali dell'item analysis (Coefficiente Alfa di Cronbach, Indice di Facilità, Punto-Biserial e altri parametri descrittivi);
- l'analisi delle criticità degli item;
- l'analisi dei punteggi grezzi e stimati con il modello IRT (*Item Response Theory*) che mette in relazione la difficoltà delle domande con l'abilità degli studenti.

CARATTERISTICHE DELLA MATRICE DATI PER L'ITEM ANALYSIS

La matrice dei dati deve contenere le risposte al test di *almeno* una coorte di studenti (non è necessario che i dati al primo e all'ultimo anno siano rilevati con gli stessi studenti).

I dati devono essere trasmessi ad ANVUR in un unico file excel che preveda:

- una riga di intestazione che per ciascuna domanda riporti la **sezione di riferimento/area tematica** (da 1 a N).
- una riga di intestazione che riporti il **numero per ciascuna domanda** (ad esempio D1, D2, D3);
- una riga di intestazione che per ciascuna domanda riporti la **chiave di risposta** (cioè la risposta corretta per quella domanda: A, B, C, D, E);
- una riga per **ciascuno studente**, contenente:
 - un identificativo univoco ma anonimo (ad esempio un numero o un codice alfanumerico);
 - l'anno di corso dello studente (I, II, III, se opportuno distinguere i laureandi: IV);
 - l'anno della somministrazione (se si tratta di un'unica somministrazione questa colonna non è necessaria);
 - le risposte al test, codificate con le lettere dell'alfabeto (a, b, c, d, e);

Ad esempio:

Inserire la sezione di riferimento/area tematica

Inserire il numero per ciascuna domanda

Identificativo studente	Anno di corso	Anno di rilevazione	Area 1	Area 1	Area ...	Area N
			D01	D02	D...	DN
			A	C	A	B
1	I	2016	a	a	b	c
2	I	2016	b	c	b	b
3	II	2017	c	99	c	c
...	III	2017	a	b	a	b

Inserire le chiavi di risposta

Nel caso di risposte mancanti è opportuno assegnare il codice 99.

Si sottolinea che non è prevista la trasmissione di alcun dato personale o informazione che permetta l'identificazione dello studente (ad esempio matricola, codice fiscale, o altro).

ITEM ANALYSIS E FEEDBACK ANVUR

Le analisi statistiche permettono di controllare se e in quale misura le prove siano valide e attendibili e quindi in grado di misurare i livelli di conoscenza, comprensione e conoscenza applicata, definiti dal GdL relativamente a un ambito disciplinare.

Il feedback fa ricorso alle tecniche dell'*Item analysis* classica e dell'*Item Response Theory* (IRT). Attraverso tali analisi è possibile analizzare alcuni aspetti essenziali:

- **Coerenza interna del test:** la coerenza interna della prova è stata verificata attraverso il calcolo del coefficiente Alfa di Cronbach, che misura il grado con cui un test rileva un singolo fattore. I test che misurano più di un fattore hanno valori di Cronbach bassi, mentre i test omogenei tendono a valori più elevati. L'Alfa può assumere valori compresi tra 0 e 1: più i valori si avvicinano a 1 e più sarà forte la coerenza interna del test, suggerendo che la prova raccoglie informazioni solo su di un fattore o tratto. La formula per il calcolo dell'Alfa di Cronbach è la seguente:

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_i \sigma_i^2}{\sigma_{tot}^2} \right)$$

Dove n è il numero di *item*, σ_{tot}^2 è la varianza del punteggio totale e σ_i^2 la varianza dell'*item* i per l'insieme di individui in esame.

Le soglie di accettabilità dell'Alfa di Cronbach sono riportate nella *Tabella 1*.

Tabella 1: Livelli dell'indice di coerenza del test

Alfa di Cronbach	
Valori	Coerenza interna
$\leq 0,59$	Inadeguata
$< 0,60 - 0,69 >$	Sufficiente
$< 0,70 - 0,89 >$	Buona
$\geq 0,90$	Ottimo

- **Efficacia degli item:** l'efficacia dei singoli *item* può essere verificata attraverso il calcolo di alcuni parametri:

- *Indice di facilità (P)*

L'indice di facilità permette di sintetizzare il livello di facilità dell'*item* e, in generale, dell'intera prova, mettendo in relazione gli studenti che hanno risposto correttamente al quesito con tutto il gruppo. Più è ampio il gruppo degli studenti che ha risposto in maniera corretta all'*item*, tanto più questa risulterà facile.

$$F = \frac{\sum_i x_i}{N}$$

Dove x_i è il punteggio dell'*item* i (dunque $\sum_i x_i$ è pari al numero delle risposte corrette) ed N il numero di soggetti. Questo indice permette quindi di rivedere l'intera prova con l'obiettivo di rendere più difficili o più semplici alcuni quesiti in modo tale da ottenere una prova eterogenea. Esistono soglie di accettabilità di questo indice: gli *item* che ottengono un punteggio compreso tra 0,25 e 0,75 possono essere considerati generalmente accettabili (*Tabella 2*). Tuttavia l'analisi di questo indice deve essere effettuata considerando gli scopi per i quali la prova è stata costruita e la definizione dei saperi minimi a partire dalla quale è stata progettata.

Tabella 2: Livelli dell'Indice di Facilità

Indice di Facilità	
Valori	Facilità
$1 \leq P \leq 0,75$	Facile
$0,74 \leq P \leq 0,50$	Medio-facile
$0,49 \leq P \leq 0,25$	Medio-difficile
$0,24 \leq P \leq 0$	Difficile

Per verificare l'efficacia degli *item*, l'indice di facilità è una condizione necessaria, ma non sufficiente: ogni valore di facilità deve essere affiancato a quello derivante dal calcolo del Punto-Biserial (P-Bis).

- *Punto-Biserial (P-Bis)*

La discriminatività di un *item* è calcolabile attraverso la correlazione Punto-Biserial e spiega se il quesito è in grado di distinguere gli studenti più preparati da quelli che lo sono meno. Si tratta di un coefficiente che si ottiene mettendo a confronto i risultati ottenuti da tutti gli studenti che hanno risposto correttamente a un singolo *item* con i punteggi di tutti gli studenti nell'intera prova, meno quell'*item*:

$$r_{p-bis} = \frac{\bar{x}_e - \bar{x}_t}{\sigma} \sqrt{\frac{p}{1-p}}$$

Dove \bar{x}_e è la media dei punteggi al test dei soggetti che hanno dato la risposta corretta all'*item*, \bar{x}_t è la media dei punteggi al test di tutti i soggetti, σ è la deviazione standard dei punteggi di tutti i soggetti e p è la frequenza relativa delle risposte esatte: $p = (x_i / N)$.

L'indice varia da - 1 a + 1. Il valore massimo di + 1 viene raggiunto quando tutti gli studenti più competenti rispondono in maniera corretta al quesito, contro nessuna risposta corretta del gruppo dei meno preparati. Il valore minimo teorico -1 si presenta quando tutti i migliori sbagliano e tutti i peggiori rispondono in maniera corretta. Se l'indice si avvicina a 0 vuol dire che c'è una mancanza di discriminatività e che l'*item* non è in grado di distinguere realmente i due gruppi di studenti.

Come per l'indice di facilità, esistono soglie di accettabilità del Punto-Biserial (*Tabella 3*), ma anche in questo caso, l'analisi dell'indice deve essere effettuata considerando gli scopi per i quali la prova è costruita. Ad esempio ci si aspetta un valore alto all'inizio del corso di studi (I anno), mentre al termine di un percorso formativo (III anno) sarà normale ottenere valori più bassi. Al termine di un percorso universitario, infatti, tutti gli studenti dovrebbero raggiungere più o meno gli stessi livelli e, di conseguenza, ci si aspetta che le distanze tra i gruppi si riducano.

Tabella 3: Livelli dell'Indice Punto-Biseriale (P-Bis)

Indice del Punto-Biseriale (P-Bis)	
Valori	Discriminatività
P-Bis $\geq 0,20$	Buona
$0,20 > \text{P-Bis} > 0,15$	Debole
P-Bis $\leq 0,15$	Inadeguata

- *Funzionalità dei distrattori*

Per ogni *item* verrà analizzata la distribuzione delle risposte degli studenti nelle quattro o cinque alternative previste. Questa analisi permette di capire se e in quale misura i distrattori (le modalità di risposta non corrette) funzionino in maniera efficace. Una domanda a scelta multipla in cui i distrattori si rivelano efficaci è caratterizzata da un'elevata frequenza di risposte per l'alternativa corretta e da una bilanciata attrattività dei distrattori. Questo vuol dire che un distrattore funziona in maniera anomala quando attrae su di sé troppe risposte (presenta troppi elementi in comune con la risposta corretta) o al contrario troppe poche risposte (troppo poco plausibile). Anche in questo caso la lettura dei dati cambia in funzione dell'anno di corso: ad esempio, mentre al primo anno ci si può aspettare una più omogenea scelta da parte degli studenti anche delle alternative errate, al termine del percorso universitario le risposte corrette dovrebbero attirare maggiormente le preferenze degli studenti, anche a scapito di una buona distribuzione di scelte per i distrattori.

Gli indici illustrati considerano il gruppo come un singolo ed unitario insieme, ma dal momento che il gruppo è formato da studenti che possono raggiungere livelli di conoscenza, comprensione e conoscenza applicata diversi, viene esaminato il grado di omogeneità o eterogeneità del campione attraverso il calcolo del **Coefficiente di Variazione (CdV)**, indice che si ottiene dal rapporto tra due parametri di una distribuzione, la media e la deviazione standard, e si esprime in percentuale:

$$CdV = \frac{s}{\bar{x}} * 100$$

Il CdV fornisce un'informazione relativa alla dispersione dei punteggi: quanto più sarà alto tanto più i punteggi risulteranno dispersi rispetto alla media. Quando il coefficiente è inferiore a 10 si può parlare di un gruppo omogeneo, mentre un coefficiente che supera 20 indica una forte eterogeneità del gruppo. Nel caso in cui la deviazione standard dalla media sia superiore alla media stessa si ottiene un CdV superiore a 100.

Ci si attende un CdV elevato all'inizio del percorso universitario e una graduale diminuzione di questo valore negli anni successivi.

Alla relazione che presenta l'analisi dei dati sono allegare diverse tavole in xls, che riportano le informazioni rilevanti, considerando diversi livelli di dettaglio:

- a) i punteggi per studente per Area e complessivi (questi ultimi calcolati anche in base al modello IRT, se supportato dal test);
- b) dati sulle caratteristiche del test (per anno di corso), come nell'esempio in Tabella 4;
- c) dati sulle caratteristiche del test per Area (per anno di corso), come nell'esempio in Tabella 5;
- d) dati sulle caratteristiche degli item (per anno di corso), come nell'esempio in Tabella 6.

Tabella 4. Esempio di dati sulle caratteristiche del test

	Anno di corso		
	1	2	3
Alfa di Cronbach	0,844	0,919	0,938
Facilità	0,491	0,537	0,696
P-Biseriale	0,173	0,234	0,263

Tabella 5. Esempio di dati sulle caratteristiche del test per Area

Area	N. Item	Anno di corso	N. Studenti	Punteggio minimo	% studenti punt. minimo	Punteggio massimo	% studenti punt. massimo	Punteggio medio	Dev. std. dei punteggi	Punteggio medio normalizzato	CdV
<i>n</i>	<i>n</i>	1	200	0	0,80%	15	0,05%	10,225	5,747	0,341	54,865
		2	200	0	0,40%	20	0,10%	12,255	5,423	0,543	26,584
		3	200	0	0,20%	25	0,25%	15,555	5,115	0,650	19,202

Tabella 6. Esempio di dati sulle caratteristiche degli item:

Area	N. Item	Anno di corso	Facilità	P-Biseriale	Chiave	Frequenze risposte				Omissioni
						A	B	C	D	
<i>n</i>	<i>n</i>	1	0,300	0,200	a	0,300	0,346	0,081	0,221	0,053
		2	0,530	0,220	a	0,529	0,214	0,031	0,180	0,046
		3	0,748	0,258	a	0,747	0,099	0,020	0,117	0,018