

Simulazione d'esame

Margherita Calderan & Ottavia Epifania

1 Scelta multipla (12 pt.)

Le prossime domande sono a scelta multipla con 4 opzioni di risposta, dove solo una è la corretta. I punteggi vengono assegnati come segue:

- +1 risposta corretta
- 0 risposta non data/errata

1.1 Quale tra le seguenti definizioni descrive correttamente il processo di misurazione?

- a) Un'operazione che permette di assegnare etichette a oggetti, eventi o fenomeni in modo arbitrario
- b) Il processo di assegnazione di numeri o simboli a proprietà di oggetti, eventi o fenomeni seguendo regole definite e condivise
- c) La rilevazione di qualità non quantificabili attraverso simboli non numerici
- d) L'assegnazione di valori numerici esclusivamente a fenomeni fisici

Risposta corretta: B

1.2 La mediana di una distribuzione è:

- a) Il valore che ha la frequenza più alta
- b) La media aritmetica dei valori centrali
- c) Il valore che divide la distribuzione ordinata in due metà uguali
- d) Il valore che minimizza la somma degli scarti al quadrato

Risposta corretta: C

1.3 Secondo gli assiomi di Kolmogorov, la probabilità di un evento certo è:

- a) 0
- b) 0.5
- c) 1
- d) Dipende dallo spazio campionario

Risposta corretta: C

1.4 Un test che misura il costrutto che dichiara di misurare e non altro è caratterizzato da una buona:

- a) Attendibilità
- b) Validità di costrutto
- c) Validità ecologica
- d) Standardizzazione

Risposta corretta: B

1.5 La discriminatività di un item rappresenta:

- a) La proporzione di rispondenti che hanno scelto il punteggio massimo
- b) La capacità dell'item di distinguere tra rispondenti con livelli alti e bassi di tratto latente
- c) La media delle risposte all'item
- d) L'errore casuale associato all'item

Risposta corretta: B

1.6 La comunalità h_i^2 di un item rappresenta:

- a) La proporzione di varianza dell'item spiegata dai fattori latenti
- b) L'errore specifico dell'item
- c) La media dei loading
- d) La correlazione tra due item

Risposta corretta: A

1.7 Lo scree plot è uno strumento utilizzato per:

- a) Visualizzare la correlazione tra item
- b) Decidere il numero di fattori da estrarre
- c) Valutare la normalità della distribuzione
- d) Stimare la difficoltà degli item

Risposta corretta: B

1.8 Il criterio di Kaiser per determinare il numero di fattori suggerisce di:

- a) Estrarre i fattori con autovalore maggiore di 1
- b) Estrarre tanti fattori quanti sono gli item
- c) Estrarre i fattori con loading maggiore di 0.4
- d) Estrarre sempre un solo fattore

Risposta corretta: A

1.9 La tendenza a essere sistematicamente d'accordo con le affermazioni proposte dagli item è definita:

- a) Estremismo
- b) Evasività
- c) Acquiescenza
- d) Desiderabilità sociale

Risposta corretta: C

1.10 L'assunzione di indipendenza locale stabilisce che:

- a) Le variabili osservate sono indipendenti dalle variabili latenti
- b) La correlazione tra variabili osservate si riduce dopo aver controllato l'influenza della variabile latente
- c) Le variabili latenti sono sempre correlate tra loro
- d) Gli errori di misurazione sono indipendenti dalla variabile latente

Risposta corretta: B

1.11 Quale funzione di R restituisce la probabilità cumulata di un valore in una distribuzione normale?

- a) `pnorm()`
- b) `dnorm()`
- c) `qnorm()`
- d) `rnorm()`

Risposta corretta: A

1.12 Quale delle seguenti affermazioni è vera?

- a) L'EFA e la CFA si possono svolgere sullo stesso identico set di dati
- b) La CFA non può essere eseguita su item dicotomici
- c) EFA e CFA sono la stessa cosa, cambia solo la scalatura dei parametri
- d) Standardizzando la varianza a livello latente, si ottiene una saturazione per ogni singolo item

Risposta corretta: D

2 Output (19 pt. complessivi)

Per ogni domanda in questa sezione viene mostrato un **output** (tabella, grafico, output di una funzione R) e si chiede di **interpretarlo a parole** (scritte), descrivendo cosa rappresenta e cosa permette di concludere sui dati o sul modello.

2.1 Output (4 pt.)

Categoria	Freq. assolute	Freq. cumulate	?
1	4	4	0.08
2	9	13	0.18
3	11	24	0.22
4	13	37	0.26
5	13	50	0.26

Descrivere la tabella: cosa rappresentano le frequenze assolute e quelle cumulate? Quante persone hanno risposto con un valore ≤ 3 ? Qual è la moda della distribuzione? Che informazione contiene l'ultima colonna?

Le frequenze assolute rappresentano il numero di unità statistica che mostrano quel determinato livello della variabile. Nell'esempio, ci sono 4 unità statistiche che hanno risposta categoria 1. Le frequenze cumulate rappresentano il numero di unità statistiche del campione che presentano un livello uguale o inferiore a quello preso in considerazione. Il totale delle frequenze cumulate deve essere sempre uguale all'ampiezza campionaria. In questo caso, l'ampiezza campionaria è 50. Ci sono 24 persone che hanno risposto con un valore minore o uguale a 3, come si vede dalla frequenza cumulata associata alla categoria 3. La moda è la categoria di una variabile a cui è associato il valore massimo delle frequenze. In questo caso, abbiamo una distribuzione bimodale perché le categorie 4 e 5 presentano la stessa frequenza massima di 13. L'ultima colonna rappresenta le frequenze relative, ottenute dividendo le frequenze assolute associate ad ogni categoria per l'ampiezza campionaria.

2.2 4 pt

	I1	I2	I3	I4
I1	1.00	0.69	0.11	0.05
I2	0.69	1.00	0.12	0.09
I3	0.11	0.12	1.00	0.63
I4	0.05	0.09	0.63	1.00

Descrivere la struttura di correlazione. Quali coppie di item sembrano misurare lo stesso costrutto latente? Quanti fattori vi aspettereste di estrarre con un'analisi fattoriale?

Nell'output è riportata una matrice di correlazione tra 4 item. L'item 1 mostra una correlazione molto forte e positiva con l'item 2, l'item 3 mostra una correlazione molto forte e positiva con l'item 4. La correlazione tra l'item 3 e l'item 1 e 2 invece è più bassa, anche se positiva. Da questa tabella è possibile immaginare che esistano due fattori latenti che spiegano, rispettivamente, la correlazione tra I2 e I1 e quella tra I3 e I4.

2.3 4 pt.

	Media	normalizzato
I1	3.59	0.65
I2	2.23	0.31
I3	2.98	0.50

Spiegate come è stato calcolato il π normalizzato e a cosa serve la normalizzazione. Quale item risulta il più "facile"?

Il punteggio normalizzato viene calcolato per rapportare il valore osservato al minimo e massimo teorico della scala di misura, permettendo l'interpretazione della difficoltà dell'item politico su una scala da 0 a 1, come per gli item dicotomici. Anche se viene definita difficoltà, il

punteggio tra 0 e 1 è un indice di facilità dell'item (più il valore è grande, più è facile rispondere a valori a destra della scala, ovvero ai valori alti). Nel caso dell'esempio, l'item 1 è l'item più facile, ovvero l'item che mediamente riceve i punteggi più a destra della scala.

2.4 7 pt.

This is lavaan 0.6-19 -- running exploratory factor analysis

```

Estimator                      ML
Rotation method                 OBLIMIN OBLIQUE
Oblimin gamma                  0
Rotation algorithm (rstarts)   GPA (30)
Standardized metric            TRUE
Row weights                    None

Number of observations          400

```

Overview models:

```

          aic      bic      sabic  chisq df pvalu  cfi rmsea
nfactors = 2 9421.77 9525.54 9443.05 289.10 19  0.00 0.74  0.19
nfactors = 3 9156.15 9287.87 9183.16   9.49 12  0.66 1.00  0.00
nfactors = 4 9161.05 9316.72 9192.97   2.39  6  0.88 1.00  0.00

```

Eigenvalues correlation matrix:

```

  ev1   ev2   ev3   ev4   ev5   ev6   ev7   ev8   ev9
 2.36  2.01  1.97  0.55  0.50  0.44  0.41  0.38  0.37

```

Number of factors: 2

Standardized loadings: (* = significant at 1% level)

```

          f1      f2      unique.var  communalities
item1  0.73*           0.46           0.54
item2  0.80*           0.36           0.64
item3  0.74*           0.45           0.55
item4           0.76*  0.42           0.58
item5           0.80*  0.37           0.63
item6           0.74*  0.44           0.56
item7           .           1.00           0.00
item8           .           0.99           0.01
item9           .           1.00           0.00

```

	f2	f1	total
Sum of sq (obliq) loadings	1.78	1.74	3.52
Proportion of total	0.51	0.49	1.00
Proportion var	0.20	0.19	0.39
Cumulative var	0.20	0.39	0.39

Factor correlations: (* = significant at 1% level)

	f1	f2
f1	1.00	
f2	0.09	1.00

Number of factors: 3

Standardized loadings: (* = significant at 1% level)

	f1	f2	f3	unique.var	communalities
item1	0.74*		.*	0.44	0.56
item2	0.80*			0.37	0.63
item3	0.74*			0.45	0.55
item4		0.76*		0.42	0.58
item5		0.80*		0.37	0.63
item6		0.75*		0.44	0.56
item7	.		0.66*	0.57	0.43
item8	.	.	0.71*	0.48	0.52
item9			0.73*	0.46	0.54

	f2	f1	f3	total
Sum of sq (obliq) loadings	1.78	1.74	1.49	5.01
Proportion of total	0.36	0.35	0.30	1.00
Proportion var	0.20	0.19	0.17	0.56
Cumulative var	0.20	0.39	0.56	0.56

Factor correlations: (* = significant at 1% level)

	f1	f2	f3
f1	1.00		
f2	0.10	1.00	
f3	0.02	0.04	1.00

Number of factors: 4

Standardized loadings: (* = significant at 1% level)

	f1	f2	f3	f4	unique.var	communalities
item1	0.95*	*	*	*	0.09	0.91
item2	0.42*	0.35*		.*	0.53	0.47
item3		1.00*		*	0.00	1.00
item4			0.76*		0.42	0.58
item5	.		0.80*		0.36	0.64
item6	.		0.75*		0.43	0.57
item7	.			0.64*	0.57	0.43
item8		.	.	0.70*	0.49	0.51
item9	.	.*		0.75*	0.44	0.56

	f3	f4	f2	f1	total
Sum of sq (obliq) loadings	1.78	1.49	1.21	1.18	5.66
Proportion of total	0.31	0.26	0.21	0.21	1.00
Proportion var	0.20	0.17	0.13	0.13	0.63
Cumulative var	0.20	0.36	0.50	0.63	0.63

Factor correlations: (* = significant at 1% level)

	f1	f2	f3	f4
f1	1.00			
f2	0.58	1.00		
f3	0.09	0.07	1.00	
f4	-0.05	0.07	0.04	1.00

Cosa permette di fare la funzione `efa()` di `lavaan` quando si specifica `nfactors = 2:4`? In base agli indici di fit riportati, quale soluzione fattoriale appare preferibile? La soluzione fattoriale preferibile in base agli indici di fit è in linea con i principi di parsimonia e struttura semplice? Che principio non si può evincere dall'output riportato?

La funzione `efa()` del pacchetto `lavaan` permette di stimare diversi modelli di analisi fattoriale esplorativa. Il numero di fattori da estrarre viene specificato attraverso l'argomento `nfactors`. Nel caso in esame, la funzione stima 3 modelli di analisi fattoriale esplorativa, uno a due, uno a tre e uno a 4 fattori, in questo caso correlati come si evince dalla voce `Rotation method OBLIMIN OBLIQUE` dell'output. In base agli indici di fit comparativi basati sull'entropia AIC e BIC, il modello che presenta i valori più bassi è il modello con 3 fattori. Questi indici infatti penalizzano la funzione di verosimiglianza del modello per il numero di parametri ottenuti, per cui più è basso il valore, migliore è il rapporto tra verosimiglianza e numero di parametri. La statistica chi quadro riportata nella colonna `chisq` testa l'ipotesi nulla che la matrice di covarianza attesa in base al modello a F fattori riproduca la matrice di covarianza osservata. Secondo il p value associato a questa statistica, il modello a 2 fattori va rigettato perché p

value $< .05$ mentre i modelli a 3 e 4 fattori vengono accettati secondo questo criterio. Le statistiche CFI e RMSEA devono essere rispettivamente superiore a .90 e inferiore a .10, per cui solo il primo modello viene rigettato. Considerando tutte le statistiche insieme, il modello a 3 fattori sembra essere il migliore, proprio per il rapporto tra verosimiglianza e numero di parametri. Il CFI è un indice proprio della CFA e nel caso della EFA va interpretato comparativamente con gli altri modelli, per vedere se la fit del modello cambia a seconda del numero di fattori. Relativamente al principio di struttura semplice, solo il modello a 3 fattori presenta una saturazione sostanziale degli item sui 3 fattori diversi. Nel modello a due fattori gli ultimi 3 item non saturano su nessun fattore, mentre nel modello a 4 fattori l'item 1 presenta cross loading - saturazione sostanziale su due fattori - e il secondo fattore ha solo 2 item, mentre per il principio della struttura semplice ogni fattore deve avere almeno 3 item che saturano su di esso e gli item non devono presentare cross loadings. Relativamente al principio di parsimonia, la soluzione fattoriale deve spiegare una buona quota di varianza senza però aggiungere fattori non necessari. La soluzione a due fattori spiega il 49% della varianza (colonna total della riga Proportion Var o Cumulative Var), il modello a 3 fattori il 56% e il modello a 4 fattori il 63%. Sebbene l'aumento di varianza tra i fattori sia, in termini percentuali, lo stesso, la mancanza della struttura semplice fa propendere per il modello a 3 fattori, in accordo con le statistiche di fit. Il criterio che non si può valutare attraverso questo output è il criterio della causazione, secondo cui le risposte osservate negli item e la correlazione tra di esse siano la realizzazione delle variabili latenti. Questo principio si esplora guardando la matrice dei residui, ovvero guardando le correlazioni residue tra gli item dopo aver tolto l'influenza delle variabili latente. In una situazione dove c'è effettivamente causazione, ci si aspetta che i valori delle correlazioni siano prossimi allo 0.