

## ANALISI FATTORIALE

I calcoli che vengono fatti con l'analisi fattoriale servono soprattutto per validare dei questionari quindi è così troviamo una struttura esterna per andare a vedere gli item, cioè le domande, sono costruite nel modo giusto e sostanzialmente trova dei fattori che andremo a vedere che cosa sono. infatti l'analisi fattoriale nasce proprio in ambito medico psicologico e, alla sua nascita, come tutt'oggi, lo scopo principale o meglio l'utilizzo principale che ne viene fatto è appunto relativa allo studio della struttura dei dati delle variabili che noi prendiamo in considerazione nella nostra ricerca e soprattutto per validare proprio i questionari. È una procedura puramente matematica e questo è importante perché la maggior parte delle critiche e delle finzioni diciamo così che sono state create a posteriori dell'analisi fattoriale derivano proprio dal fatto che molti autori sopra stimavano questa procedura considerandola non più solo come una procedura matematica ma come qualcosa che andava veramente che andava di individuare le variabili, le sorgenti delle variabili che stavamo studiando, qualcosa di magico che è impossibile fare. È solo una procedura matematica. È stata sovrastimata non ci permette quindi di arrivare alla conferma di un'ipotesi, ma ma è una tecnica vantaggiosa che ci fornisce delle informazioni difficili da ottenere in un altro modo. In realtà esiste anche un'analisi fattoriale confermativa, quindi che farebbe parte del ramo della statistica confermativa, ma non ne parleremo perché è criticatissima sia io che la professoressa Benassi che il professor Bolzani non riteniamo sia possibile utilizzarla in questo senso. Molti la utilizzano ma non ha una base logica, resta. Una procedura matematica e sappiamo che se vogliamo fare una somma di tanti numeri di cellulare matematicamente corretto ma non ha alcuna logica. le procedure matematiche noi possiamo utilizzarle come ci pare ma poi dal punto di vista logico magari non hanno un senso.

Come lavora l'analisi fattoriale?

Innanzitutto utilizza variabili quantitative, gaussiane, correlate fra loro che vengono analizzate allo stesso livello. In realtà non è che noi prima di fare un'analisi fattoriale dobbiamo andare a vedere se le variabili, le domande gli item di partenza sono correlati tra loro. Le variabili in genere sono gli item di un questionario e diamo per scontato che siano correlate fra loro. In psicologia si presuppone k un qualsiasi questionario, studio, k facciamo per analizzare un argomento specifico abbia domande al suo interno che sono correlate fra loro.

Non solo se si vogliono indagare una funzione cognitiva ma anche ad esempio nel caso in cui si voglia indagare la predisposizione ad una scelta politica piuttosto che ad un'altra; ci possono essere delle domande che apparentemente sembrano non correlate tra loro, è ovvio che chiederemo l'età, che cosa fa nella vita, il livello socio culturale, ecc., in realtà queste sono correlate perché ci

servono per andare a studiare l'argomento che prendiamo in considerazione in questo momento cioè la scelta politica. Quindi si presuppone, soprattutto in psicologia, di qualsiasi studio facciamo su un argomento specifico, qualsiasi questionario per indagare un argomento specifico, le domande al suo interno siano bene o male correlate fra loro.

Le operazioni matematiche che sono proprie dell'analisi fattoriale sono:

1) innanzi tutto creo una matrice di correlazione, poi utilizzo una procedura detta estrazione di fattori ortogonali e si effettua una rotazione (non sono obbligatorie, non fanno parte dell'analisi fattoriale di base, ma sono state inserite per rispondere ad alcune critiche che sono state fatte all'analisi fattoriale)

Prende, in base alle risposte date agli item, quali sono le risposte comuni date agli item, cioè quali item indagano uno specifico argomento, dividendo in sotto argomenti.

Lo scopo principale quindi dell'analisi fattoriale è ridurre il numero delle variabili in studio, ma non dell'informazione, cioè sostanzialmente ciò che fa l'analisi fattoriale è prendere il questionario già somministrato cioè le risposte che sono state date agli item e li raggruppa in fattori che non sono altro che nuovi item. Però ogni item rappresenterà, a questo punto, 1 insieme di domande di partenza. esempio: mettiamo che il questionario abbia 50 item, l'analisi fattoriale va' a vedere, in base alle risposte che sono state date a questi item, qual è l'informazione comune a più item cioè quali sono gli item che sostanzialmente indagano specificatamente nel nostro argomento un sotto argomento che ne fa parte. per esempio nel caso indaghiamo la memoria in generale, ci saranno delle domande relative alla Works immemori, delle domande relative alla memoria a breve termine, delle domande relative alla memoria a lungo termine, e così via. Quindi ci saranno degli item . Che per forza sono raggruppati ce ne saranno cinque per ogni tipo di memoria ad esempio. lo scopo dell'analisi fattoriale è dirci quali sono i raggruppamenti che vanno a studiare una parte dell'argomento generale. l'analisi fattoriale in genere si usa per un nuovo questionario è uno strumento che serve per valli dare un nuovo questionario oppure per vedere se un questionario tradotto ad esempio dall'inglese è stato tradotto bene, nel modo giusto in termini di senso. Per esempio ci può dire se ci sono degli item, delle domande ripetitive, k ci danno la stessa informazione, dandoci così la possibilità di scartare il superfluo. In un questionario è sempre bene scartare ciò che è superfluo, soprattutto quelli lunghi che possono essere soggetti più facilmente a noia da parte di chi li deve compilare, in questo senso è bene ridurre il numero delle domande, certo non troppo, ma tenere un numero giusto di domande perché ad esempio se un questionario a 60 domande, nelle prime 30 ci impegniamo, ma poi arrivi alla trentacinquesima e ci diciamo va beh ci stiamo mettendo troppo tempo e cominciamo a tirare via.

2) trasformazione delle variabili in studio in variabili indipendenti. Abbiamo detto che gli item in generale sono domande correlate tra di loro, con questa tecnica e con l'estrazione dei nuovi fattori, che non sono altro che il raggruppamento degli item di partenza, i fattori k ne ricava saranno separati, cioè avranno una correlazione è pari a 0

Prende le domande del questionario e vede se è possibile individuare delle suddivisioni degli item per argomento. Divide in sottoargomenti. E cerca di trovare delle variabili (sottoargomenti) indipendenti o, meglio, ortogonali (significa "indipendenti l'una dall'altra"). L'estrazione dei nuovi fattori, che altro non sono se non i raggruppamenti dei vari item di partenza, è un'operazione matematica. Bisogna ricordarlo perché nella realtà ciò non sarebbe possibile.

3) individuazione delle sorgenti delle variabili. Alcuni studiosi pensano che dall'analisi dell'argomento di un questionario si possa risalire alle sorgenti delle matrici dell'argomento stesso. Cosa vuol dire? Torniamo all'esempio della memoria: Loro pensano che l'analisi fattoriale individuando mettiamo cinque sottogruppi, andando a vedere in ogni sottogruppo quali sono le domande che vengono raggruppate in ogni fattore, danno un nome ad ogni sottogruppo, un nome fittizio, ad esempio Gino. e poi diciamo la memoria è composta da Gino, Dino, Pino, ecc. ma qualsiasi nome possiamo dare a questo fattore è una nostra interpretazione e quindi Trovare le sorgenti delle variabili in linea teorica è vero, ma il fatto di dire, di interpretare che i raggruppamenti degli item siano i nuclei fondamentali che compongono l'argomento principale è una nostra interpretazione. sicuramente 10 sperimentatori che debbano fare un'analisi fattoriale sullo stesso argomento sicuramente daranno nomi diversi ai sottogruppi, ognuno raggrupperà i sotto argomenti a modo suo e tutto è frutto dell'interpretazione soggettiva, è il significato k gli diamo noi.. Con un'operazione matematica lineare è praticamente impossibile in realtà: ci dà un sacco di informazioni questo analisi, è molto utile, però non è magica.

Critiche:

- trova variabili artificiali, aleatorie, è possibile rintracciarne un numero infinito. Siamo noi che gli diamo i parametri cioè gli diciamo trova "tot" fattori. anche qui a riguardo alle sorgenti generatrici di cui parlavamo prima, se possiamo essere noi a determinarne il numero, come facciamo a poter dire che queste sono veramente le sorgenti reali? Ovviamente ci sono vari parametri per definire questo tot, ci sono vari modi logici, matematici, per decidere la numerosità dei fattori da tenere. (in realtà la prima operazione che fa è quella di creare tanti nuovi fattori (e quando parlo di fattori in questo campo, parlo di variabili viste come raggruppamenti di informazioni trasportate dagli ha item di partenza) quante sono le variabili di partenza. In 1 secondo tempo sarà lo sperimentatore a definire quante tenerne. Sono quindi variabili create da noi, create matematicamente, sono

artificiali, quindi possono rappresentare la realtà ma non sono la realtà. e possiamo trovarne un numero infinito.

Sono indipendenti tra loro, per questo non riflettono la realtà psicologica dove è difficile trovare fenomeni non correlati.

Ruolo indiscusso di questa tecnica: validazione di questionari e riduzione del numero delle variabili in studio, che così non sono più correlate fra loro.

vantaggi principali: Riduzione delle variabili, ma non delle informazioni. quindi possiamo con questa tecnica ridurre il numero delle domande, degli item del questionario, eliminando quelle che sono costruite male, che sono ridondanti o ininfluenti, senza perdere le informazioni importanti, quelle che volevamo ottenere. crea dei fattori che rappresentano la stessa realtà del questionario, dei dati che abbiamo raccolto, ma che sono indipendenti tra di loro, e per alcuni versi questo è un vantaggio, e non va a vedere le differenze tra le diverse variabili. Gli item ma vengono considerati tutti allo stesso livello da “quanti anni hai” fino a “quanto sei d'accordo con questa cosa”, “cosa ne pensi di questa cosa”. Non fa delle differenze di importanza tra le variabili, ci dice solo se sono costruite bene.

La prima operazione che viene fatta dall'analisi fattoriale è:

Standardizzazione dei dati raccolti:

i punteggi grezzi raccolti con il questionario vengono trasformati in punti z, in modo che possano essere confrontati fra di loro. Formula: media dei soggetti meno la media generale, diviso la deviazione standard. In questo modo possiamo non solo confrontare i dati grezzi fra di loro, ma anche confrontare vari questionari tra di loro, le scale di diversi questionari tra di loro o anche nello stesso questionario ci possono essere domande dicotomiche, scale Likert a 3, 5 o 7 intervalli che possono essere così confrontate tra di loro.

Una variabile gaussiana è detta z quando ha media 0 e varianza 1

trasformiamo quindi tutti i punteggi in altri punteggi che hanno tutti i media zero a varianza uno. Perché standardizzare i dati raccolti? Perché semplifica i conti. come dicevo prima ci permette di confrontare i diversi questionari, le diverse scale di misurazione.

cos'è la varianza? la possiamo considerare come la quantità di informazione che viene trasportata dagli item. perché in realtà l'informazione che ci deriva da qualsiasi studio stiamo facendo è data dalla differenza che troviamo in un soggetto rispetto ad un altro. L'informazione di un questionario è data proprio dalle informazioni che raccolgo su ciò che distingue i soggetti. Anche x' le H sperimentali sono sempre ipotesi di differenza. Ciò che fa l'analisi fattoriale è la compensazione della varianza, quindi delle informazioni, in nuove variabili fittizie dette fattori, senza la perdita di informazioni che mi sono utili. Quindi maggiore è la variabilità e maggiore è l'informazione. ciò è

maggior nella variabilità delle risposte che ottengono il mio questionario e maggiore è l'informazione che ottengo.

Con l'analisi fattoriale, sostanzialmente, la 1° operazione k fa con l'estrazione dei fattori è la creazione di nuove variabili che sono i raggruppamenti delle variabili di partenza, o meglio i raggruppamenti delle informazioni trasportate dalle variabili di partenza, intendendo gli item, le domande, i dati che io sono riuscita a cogliere con il mio questionario. E all'inizio ne crea tante quante sono le variabili di partenza, gli item di partenza.

La varianza complessiva all'inizio, cioè la quantità totale in informazione trasportata dai nuovi fattori, risulta ora uguale al numero degli item di partenza. perché, come dicevo prima, io ho standardizzato i dati, ed avendo ogni variabile media zero e varianza, se per esempio 12 variabili avrò varianza uguale ad 12. Siccome io creano un numero di variabili nuove pari al numero di partenza, la varianza sarà sempre 12.

La 1° operazione, dopo la standardizzazione dei dati, è la creazione della matrice di correlazione (R) che altro non è se non una tabella che ci dice come correlano, in questo caso, gli item di partenza e quindi le domande del mio questionario, tra di loro. Ovviamente parlo dei dati raccolti con il mio questionario, non è che si correlano le domande, si correlano i risultati.

La matrice di correlazione, sostanzialmente correla ogni item con se stesso e con tutti gli altri. La prima matrice di correlazione viene chiamata matrice R e, appunto, come vi dicevo, mostra tutte le relazioni possibili tra le variabili di partenza. Le variabili vengono correlate a 2 a 2. correla e il primo item con se stesso, il primo item col secondo, il primo col terzo, e così via.

quando sono correlate due variabili? 2 variabili sono statisticamente correlate fra loro quando al variare di 1 varia anche l'altra. La correlazione può essere positiva o negativa. Non è detto che la correlazione sia diretta o stabile, l'importante è che una influenza l'altra. l'indice di correlazione ci dà proprio l'idea di quanto ho una sia correlata con l'altra, di quanto una influenzi l'altra. La correlazione =1 si trova solo quando si correla 1 item con se stesso perché matematicamente l'uguaglianza non esiste ed è proprio per questo che quando facciamo una ricerca dobbiamo sempre cercare delle differenze. È un problema che spesso viene fuori anche dai ricercatori, non si può dimostrare l'uguaglianza, bisogna sempre porre l'ipotesi come differenza.

Avremo un n° di nuove variabili che vengono chiamate fattori corrispondenti al numero delle variabili iniziali ed avremo una varianza complessiva uguale al numero dei fattori iniziali perché ogni variabile ha media = 0 e varianza = 1.

La seconda operazione che viene fatta è l'estrazione dei fattori. ci sono vari metodi per estrarre i fattori. Ancora una volta vi ripeto che cosa sono i fattori: dai dati raccolti in partenza da, guardo qual è l'informazione in comune tra gli item e guardo se posso dividere questa informazione

comune in sottocategorie e per fare questo, estraggo questa informazione comune, per questo motivo l'operazione si dice estrazione dei fattori. In realtà è la creazione di nuovi fattori ma si dice estrazione proprio perché cerco di estrarre informazione comune e di metterla in nuove variabili..

Metodi di estrazione più usati sono:

metodo delle componenti principali

metodo della max verosimiglianza

metodo dei minimi quadrati.

In realtà ne esistono una decina, e tutti funzionano in modo leggermente diverso, cercano di rispondere ad alcune critiche piuttosto che ad altre, ma quello più usato in assoluto è quello delle componenti principali (che secondo la professoressa Giovagnoli non è neanche il più utile, il più efficace) o anche quello della massima verosimiglianza (da lei preferito).

### Metodo delle Componenti Principali

Tale metodo ci permette di creare delle variabili artificiali, dette fattori, fra loro ortogonali, cioè indipendenti e che abbiano correlazione = 0. non sono altro che un'altra correlazione mai fattori si pensa proprio che non si influenzino, o meglio, non è una questione di influenzamento, cioè dicono che non trasportano informazioni che sono collegabili tra di loro. quindi non c'è nessun tipo di relazione negativa o positiva, correlazione è pari a zero.

cosa sono i fattori?

I fattori sono combinazioni lineari delle variabili sperimentali, si ottengono cioè dalla somma dei prodotti delle singole variabili sperimentali, o meglio dei loro valori z, per gli opportuni coefficienti. Il valore e il segno di questi coefficienti indicano quanto e come il singolo fattore sia legato alle  $\neq$  variabili sperimentali. Ovvero il valore del coefficiente che noi troviamo ci fa vedere quanto gli item che un fattore cerca di raggruppare, quanto l'informazione che il fattore cerca di raggruppare in se stessa è distribuita nei vari item di partenza. cioè si inizia il primo strattone a una quantità di informazione parla il 20% dell'informazione totale nel nostro questionario.

Come facciamo a sapere quali sono item che sono stati raccolti in questo primo nucleo?

C'è un'altra matrice che correla il fattore e i vari item, nella quale possiamo vedere quali sono gli item più correlati con il primo fattore; qual è l'informazione trasportata dai vari item che io raggruppo nel primo fattore. E c'è un coefficiente che è un numero che ci dice quanta parte dell'informazione totale è trasportata da quel fattore. I fattori (che io estraggo) non avranno più varianza = 1. Ricordiamo che la varianza è la quantità di informazione, e siccome io vado a diminuire gli item di partenza e li vado a raggruppare in dei nuclei, è ovvio che l'informazione sarà raggruppata in questi nuclei e di conseguenza anche la varianza verrà raggruppata in questi nuclei.

Il metodo delle componenti principali utilizza un processo a cascata per cui il 1° fattore estratto spiega il max della varianza ed è il + importante. Cioè il primo fattore raggruppa la maggior parte dell'informazione comune di tutto il questionario. Questo metodo cerca di mettere in comune tutta l'informazione in comune e cerca di concentrarla nel 1° fattore, guarda cosa rimane, in questa rimanenza guarda cosa c'è in comune e lo mette nel 2° fattore, e così via. Per cui il 1° fattore ha il > parte dell'info e gli altri fattori a cascata ne avranno sempre meno fino ad arrivare a fattori con varianza minore di uno, che vanno scartati perché hanno una quantità di informazione inferiore degli item di partenza.

La tecnica matematica, cioè questa combinazione lineare che viene utilizzata per estrarre i fattori col metodo delle componenti principali: Si avvale del calcolo degli autovalori e degli autovettori.

Gli autovettori sono i fattori, quindi le nostre nuove variabili estratte, e l'autovalore del fattore è la quantità di informazione che viene trasportata.

Vengono calcolati direttamente dalla matrice di correlazione R attraverso un processo algebrico.

per estrarre i fattori con questo metodo calcola autovalori e autovettori, che sono combinazioni lineari, quindi è un'operazione matematica come vi ho detto fin dall'inizio, algebriche delle variabili sperimentali di partenza e crea questi autovalori e autovettori che non sono altro che le nostre nuove variabili viste come raggruppamento degli item di partenza e la quantità di informazione trasportata dalle nostre nuove variabili.

Autovalori = quantità di varianza di un fattore e viene detta anche comunalità, perché raggruppa tutto ciò che c'è in comune negli item di partenza.

Autovettori = sono i fattori o componenti

Formula:  $R$  per autovettore è uguale ad autovettore per autovalore.

L'Autovettore moltiplicato per la matrice di correlazione  $R$  non sarà altro che se stesso a meno di una costante che è il suo autovalore.

esempio su SPSS:

dati relativi a un questionario sul mobbing composto da 43 item.

in questo caso in realtà le variabili non sono gaussiane perché è stata utilizzata una scala Likert a 5 punti. Però guardando i dati vediamo che si avvicinano molto ad una curva normale e dunque la consideriamo gaussiana. Mancando dei dati considera solo 41 variabili. Nella matrice di correlazione, sulla base di essa, estrae i nuovi fattori artificiali che corrispondono allo stesso numero delle variabili di partenza, poi ne estrae 12 perché i primi 12 hanno varianza  $> 1$ , mentre i successivi hanno una varianza  $< 1$  e quindi non li considera. Praticamente SPSS, se non specifichiamo diversamente, ci consiglia di tenere solo i fattori con varianza non inferiore a 1, e cioè che trasportano una quantità di informazione non inferiore a quella che veniva trasportata in

origine dall'item di partenza. In questo caso ci consiglia di tenere solo i primi 12 fattori che trasportano una quantità di informazione pari al 63% del totale.

Di solito 63 – 64% è un indice standard che viene ritenuto come quantità sufficiente di informazione. È una buonissima percentuale cioè ci permette di perdere una quantità di informazione che non è molto importante per la nostra ricerca, per ciò che si vuole studiare.

Prima vi dicevo che utilizza questo meccanismo cascata per cui il primo fattore avrà una percentuale di informazione, il secondo meno, il terzo meno, e così via fino all'ultimo che avrà una varianza, quindi una quantità di informazione trasportata, pari a uno e cioè uguale alla quantità di informazione trasportata da un item di partenza. dal tredicesimo in giù la varianza è inferiore a 1 e ciò sta a significare che sono informazioni in più o magari già raccolte in altri modi.

Lo scree plot è un grafico che ci dà visivamente l'informazione sui vari fattori.

Molti studiosi si basano su questo grafico per decidere quanti fattori tenere xchè a a volte ci sono degli scalini tra un fattore dell'altro evidenti. La quantità di fattori da tenere è determinata o dalla percentuale totale di informazione che si vuole tenere (es tenere tanti fattori per avere il 70% dell'info totale). Oppure si può decidere di tenere tutti i fattori con varianza pari o superiore a uno. Oppure si può guardare lo scree plot e decidere un numero tot di fattori. Se voglio diminuire o aumentare il numero dei fattori da tenere non devo ripetere l'analisi perché essa non cambia. Con altri metodi di estrazione invece bisognerebbe rifare l'analisi perché il numero dei fattori è influente sui risultati dell'analisi stessa xchè modifica gli equilibri dell'informazione trasportata.

Guardando la matrice e guardando quanto ogni fattore estratto correla con gli item di partenza, si riesce a capire quanta informazione è trasportata da ciascun fattore. Posso risalire agli item raccolti in ogni fattore, guardare che cosa le domande sostanzialmente vogliono sapere e dare un nome a questo fattore. (es il 1 ° fattore raggruppa 25 domande e gli do un nome, il 2° fattore raggruppa 10 domande e gli do un nome). Ovviamente il primo fattore essendo il più correlato con gli item di partenza raggrupperà l'argomento principale. Sono nomi che io gli do e sono fittizi. Quando si tratta di argomenti complessi come in psicologia dare il nome a questi fattori è molto rischioso perché ciò che secondo me alcune domande vanno ad indagare non è detto che sia condiviso anche dagli altri, che gli altri pensino la stessa cosa. È utile per creare delle sottoscale e soprattutto per scartare quegli item che magari non vi servono o perché non mi dà informazioni utili o perché la domanda non è espressa bene o perché la stessa informazione mi viene data anche da altre domande e quindi posso decidere di ridurre.

La quantità di varianza riportata da ogni fattore è l'autovalore.

un'altra cautela che devo osservare è quella di ricordare sempre che queste sono operazioni matematiche che io applico in ambito psicologico e quindi devo stare attento.



Non importa che impariate esattamente come vengono calcolati gli autovettori e gli autovalori, è un'operazione puramente matematica. C'è quella formula che abbiamo visto prima che indica esattamente come essi vengono calcolati. Sappiate che gli autovettori e gli autovalori vengono estratti dalla prima matrice di correlazione R, che è quella che correla gli item tra di loro, e che gli autovalori sono la quantità di informazione trasportata dagli autovettori che sono i nostri fattori.

A cosa servono i nuovi fattori, le nuove variabili?

c'è una compattazione dell'informazione, per una parte accettabile di informazioni che non era un re, ed ho un'interpretazione delle variabili in studio, posso vedere il legame che viene a crearsi tra fattore e variabili di partenza (osservando la matrice per colonna), guardando invece la matrice per riga vedo quanto ogni fattore è correlato con tutte le domande di partenza (varianza inferiore a 1), andare a vedere a quale domande si riferiscono, se sono importanti e quindi quali posso scartare.

L'analisi fattoriale sempre molto stessa per archiviare i dati di una ricerca, senza conservare tutto il materiale cartaceo che sarebbe molto ingombrante.

Sappiate che è possibile calcolare soggetto per soggetto il loro valore, cioè per ogni soggetto calcolare qual è il loro coefficiente ad ogni fattore. Come si fa?

Tali punteggi saranno variabili z con media 0 e varianza 1.

Si prende il punteggio grezzo del soggetto 1 all'item 1 e lo si moltiplica per il coefficiente di correlazione tra item 1 e fattore 1. quindi la prima risposta al primo item del soggetto uno e a molti ego per il coefficiente di correlazione che trovo correlando l'item uno con il fattore uno.

Faccio la stessa cosa per il primo soggetto con tutti gli item (1,2,3..) sommo tutti i prodotti ed ottengo il punteggio, il coefficiente del soggetto 1 al fattore 1. Questo poi lo faccio per tutti gli altri fattori e sostanzialmente mettendo via questo calcolo e la mia analisi, ho il materiale, tutti i dati sufficienti per riprendere un domani una ricerca che ho archiviato.

Avremo tanti valori soggetto per soggetto per tutti i fattori.

### Metodo dei minimi quadrati

Minimizza la somma dei quadrati degli scarti fra i dati osservati e la matrice di correlazione prodotta dal modello. Non lo spiega perché viene utilizzato molto meno.

### Metodo della max verosimiglianza

È un metodo molto usato. (secondo la prof è il più utile di tutti) Inizia e parte dalle componenti principali e poi modifica un po' le cose. Mentre lo scopo principale del metodo delle componenti principali era quello di compattare i dati, Lo scopo di questo metodo è cercare di rappresentare al meglio la realtà che stiamo studiando. Non utilizza più un metodo a cascata, ma utilizza un metodo

in cui l'informazione raccolta nei nostri fattori è ben distribuita. Cerca di andare a vedere le relazioni reali tra le variabili e l'informazione trasportata e di trasformarle nei nostri fattori. Mi spiego meglio. Estrae i nostri fattori, crea queste nuove variabili artificiali, cercando di distribuire l'informazione trasportata dai  $\neq$  fattori in modo più simile possibile alla realtà. Cioè, non è come nel modello precedente, dove si prende la maggior parte dell'informazione e la si concentra nel 1° fattore, poi si guarda quanta informazione è rimasta e la si concentra nel secondo fattore, di così via. Qui invece guarda quanta informazione comune c'è, ma che è molto legata, e la mette nel primo fattore, che non deve riferirsi a l'informazione comune, ma si concentra sul legame, sul legame più forte. Quindi cerca di rappresentare al meglio la realtà che stiamo studiando e ovviamente la distribuzione sarà più equa e non è detto che la quantità di informazione trovata alla fine sia uguale a quella trovata dalle componenti principali. cioè parte dalle componenti principali, estranei un numero di fattori uguali al numero delle componenti di partenza, però poi comincia a modificare la quantità di informazione trasportata dai fattori: prende la quantità di informazione concentrata nel primo fattore, ne prende un po' e la mette nel secondo fattore e così via. e Cerca di uniformare la quantità di informazione. Tutte le volte che modifica la quantità di varianza trasportata da un fattore fa un'analisi del  $\chi^2$  che mi dice se la soluzione nuova trovando modificando la quantità di informazione trasformata dai vari fattori è migliore di quella precedente, se rappresenta meglio la realtà che sto studiando e lo fa confrontando i dati grezzi ai fattori nuovi.

Questa tecnica: affronta la casualità in termini inversi rispetto alla probabilità: parte dai dati sperimentali e si chiede che probabilità c'è di avere una distribuzione del fenomeno di un certo tipo. Quindi, sostanzialmente, agisce in modo inverso: invece di fare 1 ipotesi e andare a vedere se c'è differenza tra i 2 gruppi, dice "io ho una distribuzione di questo tipo. Che probabilità c'è che rispecchi la realtà?" Lavora per approssimazioni successive, cioè tutte le volte che cambia la quantità di info trasportata da i fattori fa un  $\chi^2$  e confronta la realtà che stiamo studiando (quindi i dati grezzi) con i nuovi fattori estratti e dice se la soluzione che io ho trovato, in cui ho modificato la quantità di informazione, rispecchia meglio della precedente, la realtà che sto studiando. Fa un confronto, vede se va meglio, e se sì, tiene questa situazione, poi prova un'altra modifica, rifa il confronto, vede se va meglio e continua finchè non c'è più nessun miglioramento, finchè la modifica che apporta, dal confronto con la situazione precedente non comporta miglioramenti. Finchè non considera di avere ottenuto il numero di fattori e la varianza ottimale, togliendo una percentuale di informazione ininfluenza, che rappresenta nel modo più efficace la realtà in studio. Quindi lo scopo non è più la riduzione dei dati, bensì la rappresentazione migliore della realtà e poi questo ci porta anche a ridurre il numero degli item, però lo scopo principale è la rappresentazione ottimale della realtà.

Quindi lavora per approssimazioni e stima una matrice di correlazione e un'insieme di varianze che rappresentano i dati sperimentali, eliminando la ridondanza con la minima dispersione d'informazione.

La ridondanza non è altro che la ripetizione di informazione.

Lo scopo non è più la riduzione dei dati a tutti i costi, ma una rappresentazione migliore della realtà che alla fine ci porta ad una riduzione dei dati, ma non è lo scopo principale.

Ci da i risultati e poi lascia a noi la decisione di quali item da tenere. Magari ci possono essere delle informazioni, di poco conto come relazione con gli altri item (come ad es. i dati socioanagrafici), ma che per me sono importanti per determinati motivi. Quindi è lo sperimentatore che interpreta i dati e prende le decisioni.

L'analisi fattoriale è utilizzata per studiare modelli che rappresentino al meglio la realtà, caratterizzati da legami tra item e fattori, in altre parole modelli a variabili artificiali, latenti.

La tecnica della massima verosimiglianza ci offre la miglior rappresentazione della realtà possibile e ci da anche la misura di quanto bene riesca a rappresentarla.

Perché tutte le volte fa un'analisi del  $\chi^2$  fino a quando non trova il numero e la combinazione di fattori ottimale e in questo caso si dice anche che il  $\chi^2$  è significativo, cioè che il rapporto  $\chi^2$  tra i dati grezzi e la realtà rappresentata dal nostro test è ottimale.

Tale tecnica parte dall'estrazione dei fattori delle Componenti Principali.

Vengono modificati i fattori per rappresentare al meglio gli item sperimentali (quindi la quantità di informazione trasportata dai fattori).

Aumenta e diminuisce la varianza dei fattori e contemporaneamente vengono modificati anche gli altri fattori (cioè se modifica la quantità di informazione trasportata dal 1° fattore, è ovvio che viene modificata di conseguenza anche la quantità di informazione trasportata da tutti gli altri. Quindi tutte le volte la rifà l'analisi)

Ogni volta che si fa una modifica viene applicato il test del  $\chi^2$  per misurare la bontà dell'adattamento dei dati. La significatività indica che la modifica apportata è significativamente migliore rispetto ai fattori delle componenti principali (e della modifica precedente)

Il valore reale (dati grezzi) è  $\neq$  dal valore dei fattori (calcolato): infatti il valore è diverso, non è più un dato grezzo. Così il  $\chi^2$  confronta le 2 situazioni relative agli stessi dati e ci dice quanto è reale la rappresentazione attuale. Cerca di trovare un modello che rappresenti la realtà.

Otengo anche in questo caso delle variabili che tra loro non sono correlate, sono ortogonali, ma non avrò più il 1° fattore con la maggior quantità di informazione spiegata e i fattori successivi con varianza man mano sempre minore, si ha una distribuzione più equa, più simile alla realtà dell'informazione. Per creare fattori non ridondanti non è necessario creare un fattore principale con

effetto a cascata, anzi, la maggior parte dei questionari distribuisce in modo equo l'informazione tra le varie domande, tra i diversi item.

Confronto tra i modelli delle componenti principali e della max verosimiglianza:

Comp princ:

calcola i fattori attraverso una formula diretta

offre la miglior compensazione di varianza con la minima dispersione di informazione (questo non significa che rappresenti la realtà nel miglior modo)

al variare del numero di fattori da tenere l'analisi non cambia

Massima verosimiglianza:

usa la tecnica delle approssimazioni successive

trova i fattori che rappresentano al meglio la realtà

(in spss) non lavora con troppe variabili o troppo poche variabili (è un limite di spss e dell'analisi fattoriale in genere: quando ci sono troppe o troppo poche variabili l'analisi non funziona perché nel 1° caso non riesce a rappresentare bene la realtà e nel 2° caso la quantità di modifiche che deve fare è talmente elevata che richiederebbe una quantità di tempo considerevole anche per un computer)

al variare del numero dei fattori, devo rifare l'analisi

maggiore sarà il numero delle variabili di partenza e maggiore sarà la differenza dei risultati delle 2 analisi.