

Lezione 9

La regressione multipla: Approcci analitici

Argomenti della lezione:

- **Regressione standard**
- **Regressione gerarchica**
- **Regressione statistica**

Strategie Analitiche per la regressione

Tre principali strategie analitiche per l'analisi della regressione multipla. Rispondono a domande o esigenze di ricerca differenti

Regressione standard:

- ➔ **Quale è l'entità della relazione complessiva tra la VD e le VI?**
- ➔ **Quale è il contributo unico di ciascuna VI nel determinare questa relazione?**

Regressione gerarchica

Se la variabile X_1 è inserita dopo la variabile X_2 , quale è il contributo aggiuntivo che essa fornisce alla spiegazione della VD ?

Regressione statistica

Quale è la migliore combinazione lineare di VI per predire la VD in un determinato campione ?

La regressione standard

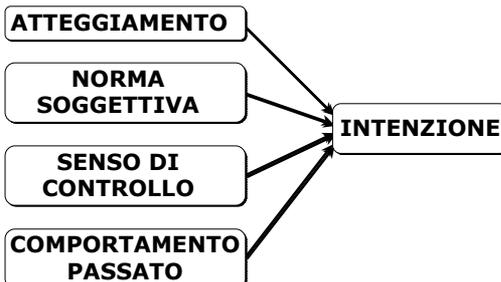
Tutte le VI vengono inserite nell'equazione allo stesso tempo

Ognuna è trattata come se fosse inserita nell'equazione dopo aver preso in considerazione tutte le altre VI

→ Ogni VI è valutata per quanto aggiunge, nella spiegazione della VD, rispetto a quanto viene spiegato da tutte le altre VI

→ Ad ogni VI corrisponde, nella spiegazione della VD, solo quella parte di variabilità che essa condivide unicamente con la VD

La variabilità che la VD condivide simultaneamente con più VI viene ad aggiungersi al R^2 ma non è assegnata individualmente a nessuna delle VI



Var	B	Beta	sr	pr	t(55)	p
Att	.15	.26	.22	.34	2.69	.01
NS	2.08	.51	.43	.58	5.32	.00
SdC	.00	.00	.00	.00	.03	.98
CP	.49	.24	.21	.23	2.58	.01
Int.	-2.27					.49

$R = .80$; $R^2 = .64$
 $F(4,55) = 24.61$ $p < .0001$

Nella regressione standard sr^2 rappresenta il contributo unico della VI all' R^2

La differenza tra la somma degli sr^2 e l' R^2 può essere considerata come la proporzione di varianza della VD spiegata simultaneamente da più VI

Nella tabella la somma degli sr^2 è:

$$(.22)^2 + (.43)^2 + (.00)^2 + (.21)^2 = .27$$

Poiché l' R^2 è uguale a .64, la differenza tra .64 e .27 è .37, ovvero il 37% della varianza della VD è spiegato simultaneamente dalle VI, ma non viene attribuito a nessuna VI in particolare

La regressione gerarchica

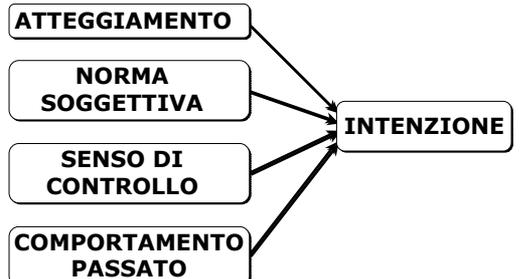
- Nella regressione gerarchica le VI vengono inserite nell'equazione secondo un ordine specificato dal ricercatore
- L'ordine di "entrata" viene assegnato secondo considerazioni teoriche o logiche

Ogni VI è valutata per quanto aggiunge, nella spiegazione della VD, rispetto a quanto è stato spiegato dalle variabili inserite precedentemente

L'analisi procede attraverso "passi" sequenziali

Si ha una partizione ordinata della varianza della VD spiegata dalle VI le quali, ad ogni passo, possono aggiungere una porzione di variabilità spiegata in più rispetto a quella sino ad allora spiegata

Il contributo di una VI può variare se la sua posizione nella gerarchia viene cambiata (cioè se viene fatta entrare in equazione prima o dopo)



Var	R	R ²	R ² C	F	p
Att	.56	.31	.31	26.49	.00
NS	.77	.60	.29	39.66	.00
SdC	.77	.60	.00	.43	.52
CP	.80	.64	.04	6.65	.01

R = .80 R² = .64
F(4,55) = 24.61 p < .0001

Nella regressione gerarchica sr² è interpretato come la quantità di varianza aggiunta all' R² da ciascuna VI nel punto in cui la VI entra nell'equazione
(spesso si definisce questa quantità come "incremental sr²". Nella tabella è indicato come R²C)

La domanda che il ricercatore si pone è la seguente:

Quanta varianza una certa VI aggiunge a R² dopo che le VI che hanno priorità maggiore sono state già considerate ?

Nell'analisi gerarchica la somma degli sr² è esattamente uguale al valore dell' R²

La regressione "statistica"

L'ordine di ingresso delle VI nell'equazione di regressione è determinato solamente da criteri statistici (di solito il valore del coefficiente sr²)

Sui medesimi criteri è basata la decisione riguardo quali variabili vengono incluse e quali escluse dall'equazione

Differenze marginali rispetto a questi criteri possono determinare conseguenze drastiche nell'importanza data alle diverse VI

Versioni della regressione statistica

Regressione forward

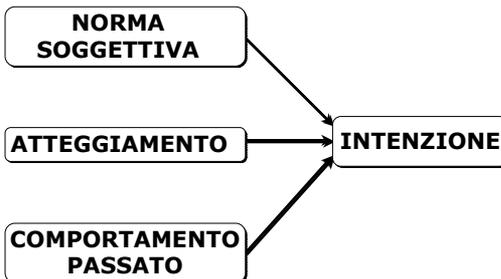
l'equazione inizialmente è "vuota".
Vengono aggiunte di volta in volta le VI che presentano la correlazione più elevata con la VD: una volta che una VI entra in equazione, vi rimane

Regressione backward

L'equazione inizialmente comprende tutte le VI in analisi: ad ogni passaggio viene eliminata la VI che non contribuisce sufficientemente alla spiegazione della VD. Una volta uscita dall'equazione la variabile non può più rientrarvi

La regressione stepwise

rappresenta un compromesso: come nella regressione forward nessuna VI è inizialmente inclusa nell'equazione e vengono aggiunte una alla volta le VI che soddisfano il criterio statistico.
Ad ogni passaggio possono essere eliminate le VI che non contribuiscono più significativamente



Step	R	R ²	R ² C	F	p
1	.72	.52	.52	63.36	.00
2	.77	.60	.08	10.29	.00
3	.80	.64	.04	7.25	.01

R = .80 **R² = .64**
F(3,56) = 33.41 **p < .0001**

Var	B	Beta	sr	pr	t	p
Step 1 (gdl = 58)						
NS	2.97	.72	.72	.72	7.96	.00
Step 2 (gdl = 57)						
Att	.17	.30	.27	.27	3.21	.00
NS	2.42	.59	.53	.53	5.30	.00
Step 3 (gdl = 56)						
Att	.15	.26	.23	.23	2.83	.01
NS	2.08	.51	.43	.43	5.39	.00
CP	.50	.24	.22	.22	2.69	.01

Critiche alla regressione statistica

1. Numero di gradi di libertà errato
- I risultati sono poco replicabili
 - Può non identificare l'insieme migliore di variabili indipendenti

Questi problemi sono tanto più seri:

- quanto più piccolo è il campione
- quanto maggiore è il numero dei passi effettuati
- quante più variabili indipendenti vengono considerate

Differenti metodi ⇒ Differenti risultati

- ↻ Standard ⇒ 37% di varianza non attribuibile a nessuna variabile
- ↻ Gerarchica ⇒ Atteggiamento e Norma Soggettiva spiegano la stessa quantità di varianza
- ↻ Stepwise ⇒ Norma Soggettiva variabile più importante

Sceita fra i differenti metodi

- La regressione standard è la strategia analitica migliore per studi esplorativi
- La regressione gerarchica è utile in un contesto confermativo
- La regressione statistica può risultare utile per scopi "pratici"

CONCLUSIONE

- Tecnica flessibile per studiare la relazione di dipendenza tra variabili
- Possibilità di definire modelli a priori (nel caso della regressione gerarchica)

CONCLUSIONE

Limiti della regressione

- Assunzione di assenza di errore
- Problema della multicollinearità

CONCLUSIONE

Limiti della regressione

- **impossibile considerare simultaneamente più di una variabile dipendente**
- **risultati soggetti ad interpretazioni assai differenti**

CONCLUSIONE

Importante:

- **Accertare le condizioni di applicabilità**
- **Scegliere l'approccio più adeguato per gli scopi del ricercatore**