

Psicometria con Laboratorio di SPSS 2

Regressione e anova: addendum
(v. 1.7a, 8 novembre 2022)

Germano Rossi¹
germano.rossi@unimib.it

¹Dipartimento di Psicologia, Università di Milano-Bicocca

2017-18

Variabili *dummy*

- Abbiamo visto che la regressione implica due o più variabili quantitative
- Tuttavia si possono usare come variabili indipendenti anche variabili dicotomiche (sfruttando le loro proprietà specifiche)
- E con una particolare strategia si possono trasformare le variabili categoriali in un insieme di variabili dicotomiche
- Una variabile dicotomica viene codificata con i valori 0 e 1

$$Y = b_0 + b_1X_1 + b_2X_2 + e$$

Variabili *dummy*

- Ipotizzando una regressione multipla con 2 X (di cui la prima è dicotomica) la formula generale si trasforma

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2$$

- b_0 è la costante da sommare a Y (sempre)
- b_1 è il valore da sommare a Y quando la dicotomica vale 1
- quando la dicotomica vale 0, b_1 scompare (diventa 0)
- b_2X_2 è il valore da sommare per la seconda indipendente

$$\begin{cases} \text{se } X_1 = 1 \implies Y = b_0 + b_1 \cdot 1 + b_2X_2 + e \\ \text{se } X_1 = 0 \implies Y = b_0 + b_1 \cdot 0 + b_2X_2 + e \end{cases}$$

Variabili *dummy*

- Ipotizzando una regressione multipla con 2 X (di cui la prima è dicotomica) la formula generale si trasforma

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2$$

- b_0 è la costante da sommare a Y (sempre)
- b_1 è il valore da sommare a Y quando la dicotomica vale 1
- quando la dicotomica vale 0, b_1 scompare (diventa 0)
- b_2X_2 è il valore da sommare per la seconda indipendente

$$\begin{cases} \text{se } X_1 = 1 \implies Y = b_0 + b_11 + b_2X_2 + e \\ \text{se } X_1 = 0 \implies Y = b_0 + b_10 + b_2X_2 + e \end{cases}$$

Variabili *dummy*

- Ipotizzando una regressione multipla con 2 X (di cui la prima è dicotomica) la formula generale si trasforma

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2$$

- b_0 è la costante da sommare a Y (sempre)
- b_1 è il valore da sommare a Y quando la dicotomica vale 1
- **quando la dicotomica vale 0, b_1 scompare (diventa 0)**
- b_2X_2 è il valore da sommare per la seconda indipendente

$$\begin{cases} \text{se } X_1 = 1 \implies Y = b_0 + b_11 + b_2X_2 + e \\ \text{se } X_1 = 0 \implies Y = b_0 + b_10 + b_2X_2 + e \end{cases}$$

Variabili *dummy*

- Ipotizzando una regressione multipla con 2 X (di cui la prima è dicotomica) la formula generale si trasforma

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2$$

- b_0 è la costante da sommare a Y (sempre)
- b_1 è il valore da sommare a Y quando la dicotomica vale 1
- quando la dicotomica vale 0, b_1 scompare (diventa 0)
- b_2X_2 è il valore da sommare per la seconda indipendente

$$\begin{cases} \text{se } X_1 = 1 \implies Y = b_0 + b_1 \cdot 1 + b_2X_2 + e \\ \text{se } X_1 = 0 \implies Y = b_0 + b_1 \cdot 0 + b_2X_2 + e \end{cases}$$

Variabili *dummy*

- Una qualsiasi variabile categoriale si può trasformare in un insieme di variabili dicotomiche. Ipotizziamo di avere 4 categorie:

1	⇒	1	0	0
2	⇒	0	1	0
3	⇒	0	0	1

4	⇒	0	0	0
---	---	---	---	---

- Bastano $(c - 1)$ variabili dicotomiche (con c =numero di categorie)
- Non si utilizzano tutti i singoli valori per evitare problemi di multicollinearità
- Quanto tutte le 3 X sono zero, siamo nella 4 categoria

Esempio con 1 *dummy*

Test di effetti tra soggetti

Variabile dipendente: rse

Origine	Somma dei quadrati di tipo III	gl	Media quadratica	F	Sign.
Modello corretto	321,678 ^a	1	321,678	10,901	,001
Intercetta	275208,835	1	275208,835	9326,423	,000
genere	321,678	1	321,678	10,901	,001
Errore	9029,603	306	29,509		
Totale	326994,395	308			
Totale corretto	9351,281	307			

a. R-quadrato = ,034 (R-quadrato adattato = ,031)

ANOVA^a

Modello		Somma dei quadrati	gl	Media quadratica	F	Sign.
1	Regressione	321,678	1	321,678	10,901	,001 ^b
	Residuo	9029,603	306	29,509		
	Totale	9351,281	307			

Coefficienti^a

Modello		Coefficienti non standardizzati		Coefficienti standardizzati		Sign.
		B	Errore standard	Beta	t	
1	(Costante)	33,668	,563		59,770	,000
	sex	-2,226	,674	-,185	-3,302	,001

Anova come modello generale linearizzato

$$(\sqrt{10.901} = 3.302)$$

Anova della Regressione lineare

Regressione lineare

$$(-.185^2 = .034)$$

Regressione con 1 dicotomica

Coefficienti^a

Modello		Coefficients non standardizzati		Coefficient i standardizzati	t	Sign.
		B	Errore standard	Beta		
1	(Costante)	18,867	2,166		8,710	,000
	eta	,078	,025	,162	3,123	,002
	ae	,307	,050	,341	6,168	,000
	he	,327	,069	,261	4,737	,000
	ao	-,085	,032	-,138	-2,650	,009
	cep	1,281	,555	,118	2,308	,022

a. Variabile dipendente: rse

$$\hat{RSE} = 18.867 + .078 \times \text{Eta} + .307 \times \text{ae} + .327 \times \text{he} - .085 \times \text{ao} + 1.281 \times \text{cep} \quad \text{cep}=1$$

$$\hat{RSE} = 18.867 + .078 \times \text{Eta} + .307 \times \text{ae} + .327 \times \text{he} - .085 \times \text{ao} \quad \text{cep}=0$$

- rse=autostima
- ae=Valutaz. apparenza
- ao=Orient. apparenza
- he=Valutaz. salute
- cep=credente e praticante (*dummy*)

Regressione con 3 dicotomiche

Coefficienti^a

Modello		Coefficients non standardizzati		Coefficient i standardizzati	t	Sign.
		B	Errore standard	Beta		
1	(Costante)	12,621	1,455		8,674	,000
	mocc	,098	,043	,083	2,268	,024
	musul	5,852	1,165	,188	5,021	,000
	cep	21,115	1,027	,835	20,555	,000
	cnp	8,310	1,013	,306	8,202	,000

a. Variabile dipendente: scl

$$\begin{aligned}\hat{SCL} = & 12.621 + .098 \times \text{Mocc (per tutti)} \\ & + 5.852 \times \text{Musul (se è musulmano, altrimenti 0)} \\ & + 21.115 \times \text{Cep (se è credente e praticante, altrimenti 0)} \\ & + 8.310 \times \text{Cnp (se è credente e non praticante, altrimenti 0)}\end{aligned}$$

- scl=religiosità (Santa Clara)
- mocc=Modestia vestiti
- musul=essere musulmano
- cep=credente e praticante
- cnp=credente ma non praticante

Regressione: residui

Residui

Durbin-Watson

Diagnostiche casewise

Valori anomali oltre: 3 deviazioni standard

Tutti i casi

Continua Annulla Guida

- Con le diagnostiche dei casi, vengono mostrati quelli con punto $z \geq 3$ (o a piacere)
- Il caso 280 ha un mocc=63 mentre il modello prevede un 29.4
- Il residuo standardizzato supera $z=3$ (e questo lo rende sospetto)

Diagnostiche casewise^a

Numero di caso	Residuo standard	mocc	Valore previsto	Residuo
280	3,699	63,00000000	29,43337881	33,56662119

a. Variabile dipendente: mocc