# Le variabili binarie (dummy variables) nell'analisi di regressione

# Hanno due soli valori

0 assenza di caratteristica
1 presenza di caratteristica
(altri valori numeri sono possibili, ma questi sono i più pratici)

# 4 su maschi e 6 su femmine

| | gruppo | punteggio |
|---|---|---|
| 1 | maschi | 2 |
| 2 | maschi | 4 |
| 3 | maschi | 3 |
| 4 | maschi | 1 |
| 5 | Femmine | 4 |
| 6 | Femmine | 6 |
| 7 | Femmine | 4 |
| 8 | Femmine | 7 |
| 9 | Femmine | 4 |
| 10 | Femmine | 5 |

**Report**

punteggio

| gruppo | Media | N | Deviazione std. |
|---|---|---|---|
| maschi | 2,50 | 4 | 1,291 |
| Femmine | 5,00 | 6 | 1,265 |
| Totale | 4,00 | 10 | 1,764 |

Le femmine hanno un punteggio pari a 5,00

- Usando l'equazione di regressione con la variabile dicotomica FEM (uguale a 1 per le donne e 0 per gli uomini), possiamo predire il Punteggio.

**Coefficienti[a]**

| Modello | | Coefficienti non standardizzati | | Coefficienti standardizzati | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Errore std. | Beta | | |
| 1 | (Costante) | 2,500 | ,637 | | 3,922 | ,004 |
| | FEM | 2,500 | ,823 | ,732 | 3,038 | ,016 |

a. Variabile dipendente: punteggio

# Per una donna, Fem =1

- Punteggio previsto= media
- Y = cost + molt * X
- 2,5 + 2,5*1 = 5,0
- Per ogni donna il punteggio previsto è uguale alla media delle donne.
- Se invece fem=0 (è un uomo)
- 2,5 + 2,5*0= 2,5 Media degli uomini

**Coefficienti[a]**

| Modello | | Coefficienti non standardizzati | | Coefficienti standardizzati | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Errore std. | Beta | | |
| 1 | (Costante) | 2,500 | ,637 | | 3,922 | ,004 |
| | FEM | 2,500 | ,823 | ,732 | 3,038 | ,016 |

a. Variabile dipendente: punteggio

L'equazione predice la media
2,5 quando Fem è zero (il soggetto è un uomo)
e
5,00 quando fem = 1

# Usando la variabile dicotomica UOM (1= uomo), otteniamo questi risultati:

**Coefficienti[a]**

| Modello | | Coefficienti non standardizzati | | Coefficienti standardizzati |
|---|---|---|---|---|
| | | B | Errore std. | Beta |
| 1 | (Costante) | 5,000 | ,520 | |
| | uom | -2,500 | ,823 | -,732 |

a. Variabile dipendente: punteggio

Ossia la media (5,0-2,5*1=2,5) quando UOM=1 e 5,00 (5,0 -0) per UOM =0

# La categoria di riferimento

- È quella che non compare nella codifica
- Può essere scelta secondo l'agio di interpretazione e utilizzazione.

# Gli indicatori binari

- Sono utilizzati per le variabili categoriali
- Ogni categoria viene usata per creare un indicatore binario.
- Se k sono le categorie, servono k-1 indicatori binari
- La codifica più semplice, efficace e comprensibile è 1 per presenza del carattere e 0 per assenza.

# Esempio con la codifica della scuola

**scuola**

| | | Frequenza | Percentuale | Percentuale valida | Percentuale cumulata |
|---|---|---|---|---|---|
| Validi | 1 CFP | 128 | 20,2 | 20,2 | 20,2 |
| | 2 IPSIA | 105 | 16,5 | 16,5 | 36,7 |
| | 3 ipscom | 62 | 9,8 | 9,8 | 46,5 |
| | 4 ITC | 120 | 18,9 | 18,9 | 65,4 |
| | 5 ITI | 72 | 11,3 | 11,3 | 76,7 |
| | 6 Classico | 31 | 4,9 | 4,9 | 81,6 |
| | 7 Scientifico | 80 | 12,6 | 12,6 | 94,2 |
| | 8 Magistrali | 37 | 5,8 | 5,8 | 100,0 |
| | Totale | 635 | 100,0 | 100,0 | |

# Esempi di codifiche

| Scuola | CFP | IPSIA | IPS Commercio | Ecc ecc |
|---|---|---|---|---|
| 1 CFP | 1 | 0 | 0 | |
| 2 IPSIA | 0 | 1 | 0 | |
| 3 ipscom | 0 | 0 | 1 | |
| 4 ITC | 0 | 0 | 0 | |
| 5 ITI | 0 | 0 | 0 | |
| 6 Classico | 0 | 0 | 0 | |
| 7 Scientifico | 0 | 0 | 0 | |
| 8 Magistrali | 0 | 0 | 0 | |

# Sono possibili codifiche più generali

| Scuola | Licei | Tecnici | Professionali | Altre |
|---|---|---|---|---|
| 1 CFP | 0 | 0 | 1 | 0 |
| 2 IPSIA | 0 | 0 | 1 | 0 |
| 3 ipscom | 0 | 0 | 1 | 0 |
| 4 ITC | 0 | 1 | 0 | 0 |
| 5 ITI | 0 | 1 | 0 | 0 |
| 6 Classico | 1 | 0 | 0 | 0 |
| 7 Scientifico | 1 | 0 | 0 | 0 |
| 8 Magistrali | 0 | 0 | 0 | 1 |

# Esempio con la variabile Età

**età**

| | | Frequenza | Percentuale | Percentuale valida | Percentuale cumulata |
|---|---|---|---|---|---|
| Validi | 13 | 21 | 3,3 | 3,3 | 3,3 |
| | 14 | 515 | 81,1 | 81,1 | 84,4 |
| | 15 | 87 | 13,7 | 13,7 | 98,1 |
| | 16 | 12 | 1,9 | 1,9 | 100,0 |
| | Totale | 635 | 100,0 | 100,0 | |

Variabile misurata

g4 vocabolario

| età | Media | N | Deviazione std. |
|---|---|---|---|
| 13 | 18,905 | 21 | 5,638 |
| 14 | 18,468 | 515 | 5,163 |
| 15 | 15,368 | 87 | 4,273 |
| 16 | 17,167 | 12 | 4,407 |
| Totale | 18,033 | 635 | 5,157 |

Categoria di riferimento

# Usiamo la sintassi di SPSS per costruire le variabili binarie

**Definiamo tre nuovo variabili**

- compute tredici=0.
- compute quindici=0.
- compute sedici=0.

**Istruzioni condicondizionali**

- if età eq 13 tredici=1.
- if età eq 15 quindici=1.
- if età = 16  sedici= 1.

**Oppure ricorriamo alla finestra del menu per eseguire le stesse operazioni.**

# Risultati con la regressione

**Coefficienti[a]**

| Modello | | Coefficienti non standardizzati | | Coefficienti standardizzati | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Errore std. | Beta | | |
| 1 | (Costante) | 18,468 | ,223 | | 82,919 | ,000 |
| | tredici | ,437 | 1,125 | ,015 | ,388 | ,698 |
| | quindici | -3,100 | ,586 | -,207 | -5,292 | ,000 |
| | sedici | -,301 | 1,476 | -,034 | -,882 | ,378 |

a. Variabile dipendente: g4 vocabolario

| Totale | 18,033 |
|---|---|

La costante delle regressione è uguale alla media della categoria di riferimento…

| età | zo | za | tredi | quattor | quindic | sedic | m | a13 |
|---|---|---|---|---|---|---|---|---|
| 16 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | |
| 15 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | |
| 14 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | |
| 13 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | |
| 16 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | |
| 15 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | |
| 14 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | |
| 13 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | |

# Ricodifica nelle variabili binarie

| genere | età | zo | za | tredi | quattor | quindic | sedic |
|---|---|---|---|---|---|---|---|
| maschio | 16 | 1 | 0 | 0 | 0 | 0 | 1 |
| maschio | 15 | 1 | 0 | 0 | 0 | 1 | 0 |
| maschio | 14 | 1 | 0 | 0 | 1 | 0 | 0 |
| maschio | 13 | 1 | 0 | 1 | 0 | 0 | 0 |
| femmina | 16 | 0 | 1 | 0 | 0 | 0 | 1 |
| femmina | 15 | 0 | 1 | 0 | 0 | 1 | 0 |
| femmina | 14 | 0 | 1 | 0 | 1 | 0 | 0 |
| femmina | 13 | 0 | 1 | 1 | 0 | 0 | 1 |

# Ricodifica delle interazioni

| genere | età | z | za | tredi | quattor | quindic | sedic | ma13 | ma14 | ma15 | fe16 | fe13 | fe14 | fe15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| maschio | 16 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| maschio | 15 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| maschio | 14 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| maschio | 13 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| femmina | 16 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| femmina | 15 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| femmina | 14 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| femmina | 13 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |

# Ricodifica delle interazioni

| nere | età | zo | za | tredi | quattor | quindic | sedic | maschio13 | maschio14 | maschio15 | maschio16 | femmina13 | femmina14 | femmina15 | femmina16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| aschio | 16 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| aschio | 15 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| aschio | 14 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| aschio | 13 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| mmina | 16 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| mmina | 15 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| mmina | 14 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| mmina | 13 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |

# Ricodifica delle variabili

| genere | età | zo | za | tredi | quattor | quindic | sedic |
|---|---|---|---|---|---|---|---|
| maschio | 16 | 1 | 0 | 0 | 0 | 0 | 1 |
| maschio | 15 | 1 | 0 | 0 | 0 | 1 | 0 |
| maschio | 14 | 1 | 0 | 0 | 1 | 0 | 0 |
| maschio | 13 | 1 | 0 | 1 | 0 | 0 | 0 |
| femmina | 16 | 0 | 1 | 0 | 0 | 0 | 1 |
| femmina | 15 | 0 | 1 | 0 | 0 | 1 | 0 |
| femmina | 14 | 0 | 1 | 0 | 1 | 0 | 0 |
| femmina | 13 | 0 | 1 | 1 | 0 | 0 | 1 |

# Ricodifica delle interazioni

| genere | età | zo | za | tredi | quattor | quindic | sedic | ma13 | ma14 | ma15 | ma16 | fe13 | f14 | fe15 | fe16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| maschio | 16 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| maschio | 15 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| maschio | 14 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| maschio | 13 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| femmina | 16 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| femmina | 15 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| femmina | 14 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| femmina | 13 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |

# Uso degli indicatori nella regressione

- Il ricorso agli indicatori dicotomici nella regressione soddisfa diverse esigenze:
- (1) predizione con una variabile realmente dicotomica, come il genere
- (2) parzializzazione di alcuni effetti privi di interesse (o che si vogliono controllare) in un'equazione di regressione multipla
- (3) esame delle interazioni fra indicatori diversi

# Risultati con la regressione

**Coefficienti[a]**

| Modello | | Coefficienti non standardizzati | | Coefficienti standardizzati | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Errore std. | Beta | | |
| 1 | (Costante) | 18,468 | ,223 | | 82,919 | ,000 |
| | tredici | ,437 | 1,125 | ,015 | ,388 | ,698 |
| | quindici | -3,100 | ,586 | -,207 | -5,292 | ,000 |
| | sedici | -,301 | 1,476 | -,034 | -,882 | ,378 |

a. Variabile dipendente: g4 vocabolario

| | |
|---|---|
| Totale | 18,033 |

…quando le altre categorie sono tutte uguali a zero

# Risultati con la regressione

**Coefficienti[a]**

| Modello | | Coefficienti non standardizzati | | Coefficienti standardizzati | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Errore std. | Beta | | |
| 1 | (Costante) | 18,468 | ,223 | | 82,919 | ,000 |
| | tredici | ,437 | 1,125 | ,015 | ,388 | ,698 |
| | quindici | -3,100 | ,586 | -,207 | -5,292 | ,000 |
| | sedici | 1,301 | 1,476 | -,034 | -,882 | ,378 |

a. Variabile dipen g4 vocabolario

Il coefficiente significativo indica una differenza significativa della media del gruppo di riferimento

# Risultati con la regressione

**Coefficienti[a]**

| Modello | | Coefficienti non standardizzati | | Coefficienti standardizzati | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Errore std. | Beta | | |
| 1 | (Costante) | 18,468 | ,223 | | 82,919 | ,000 |
| | tredici | ,437 | 1,125 | ,015 | ,388 | ,698 |
| | quindici | -3,100 | ,586 | -,207 | -5,292 | ,000 |
| | sedici | -1,301 | 1,476 | -,034 | -,882 | ,378 |

a. Variabile dipendente: vocabolario

Vocabolario per 15 anni
18,468-3,100*1 = 15,368

# Aggiungiamo il genere

- compute femmina = 0.
- if genere eq 2 femmina =1.
- Oppure
- compute maschio =0.
- if genere eq 1 maschio =1.
- Il primo comando genera la variabile e le assegna il valore 0, il secondo comando la trasforma secondo una variabile già presente nel file dati di SPSS

**Coefficienti[a]**

| Modello | | Coefficienti non standardizzati | | Coefficienti standardizzati | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Errore std. | Beta | | |
| 1 | (Costante) | 18,340 | ,302 | | 60,712 | ,000 |
| | maschio | -,567 | ,410 | -,055 | -1,381 | ,168 |

a. Variabile dipendente: [illegible]ocabolario

Media delle femmine (categoria di riferimento)

# L'interazione

- Gli indicatori binari possono essere usati per rilevare l'effetto moltiplicativo ovvero di interazione  fra due categorie binarie
- Età x genere:
- Maschio 13  =1
- Femmina 13 =1
- Maschio 14 =1
- Femmina 14=1
- Maschio 15 =1
- Femmina 15=1
- Maschio 16 =1
- Femmina 16=1

# Gradi di libertà e numero di indicatori

- Non tutti questi indicatori sono utili:
- Se usiamo il genere e l'età, gli indicatori delle interazioni sono limitati
- Per il genere basta un indicatore, per l'età bastano tre e per le interazioni:
- 3X1 = 3

# Poniamo queste categorie come quelle di riferimento

- Per il genere: femmina
- Per l’età: 14 anni

Perciò gli indicatori sono

- Maschio13
- Maschio15
- maschio16

# Il calcolo

- Si può calcolare direttamente l'indicatore di interazione come prodotto di due indicatori semplici
- Maschio13 = maschio * tredici

- Vale 1 per i maschi di 13 anni
- 1 *1 =1
- Vale 0 per le femmine e per le altre età.
- 0*1 =0
- 1*0 =0
- 0*0 =0

# Esaminiamo alcune variabili

**Coefficienti[a]**

| Modello | | Coefficienti non standardizzati | | Coefficienti standardizzati | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Errore std. | Beta | | |
| 1 | (Costante) | 18,340 | ,302 | | 60,712 | ,000 |
| | maschio | -,567 | ,410 | -,055 | -1,381 | ,168 |

a. Variabile dipendente: g4 v[illegible]bolario

Media dei maschi:
18,340 - 0,567*1 = 17,773

**Report**

g4 vocabolario

| genere | Media | N |
|---|---|---|
| 1 Maschi | 17,773 | 344 |
| 2 Femmine | 18,340 | 291 |
| Totale | 18,033 | 635 |