

Analisi Multilivello con *Mplus*

LEZIONE I

Valerio Ghezzi

Dipartimento di Psicologia

Sapienza – Università di Roma

CALENDARIO DEI SEMINARI

GIORNO	AULA	Orario	ARGOMENTI
9 Maggio 2016	Aula 6	14.00 – 16.00	1) Introduzione al corso; 2) La costruzione del dataset multilivello; 3) La regressione multilivello.
16 Maggio 2016	Aula X	10.00 – 12.00	Esercitazioni e approfondimenti sulla lezione precedente.
23 Maggio 2016	Aula T	10.00 – 12.00	1) La scomposizione della variabilità attraverso i livelli; 2) I modelli path-analitici multilivello; 3) Mediazione e moderazione multilivello.
30 Maggio 2016	Aula T	10.00 – 12.00	Esercitazioni e approfondimenti sulla lezione precedente.
6 Giugno 2016	Aula X	10.00 – 12.00	1) La CFA multilivello; 2) I modelli SEM multilivello; 3) L'attendibilità nel contesto dell'analisi multilivello.

- **Concetti chiave del multilivello;**
- **Giustificare l'aggregazione;**
- **La regressione multilivello.**

CONCETTI CHIAVE DEL MULTILIVELLO

Sul concetto di “Multilivello”

I modelli multilivello (talvolta chiamati *modelli gerarchici*, *random coefficients model* o *mixed models*) servono per analizzare **dati clusterizzati che hanno una natura gerarchica**;

I dati o i soggetti (o eventuali ulteriori livelli di analisi quando essi sono > 2) **possono essere considerati clusterizzati** (clustered) o annidati (nested) entro una unità di analisi sovraordinata **quando condividono un contesto di riferimento distinguibile** (es. studenti in una stessa classe, impiegati in uno stesso ufficio);

I modelli correntemente più utilizzati in letteratura tendenzialmente sono quelli che prevedono **due o, al più, tre livelli di analisi**.

Strutture di Dati “Multilivello”

	Contesti “organizzativi”	Contesti “clinici”	Contesti “cross-cultural”	Contesti “longitudinali”
LIVELLO 4	Organizzazione	SPDC	Nazione	Scuola
LIVELLO 3	Dipartimento	Psichiatra	Regione	Classe
LIVELLO 2	Ufficio	Terapeuta	Città	Studente
LIVELLO 1	Individuo	Individuo	Individuo	T0,T1,T2,T3

Sulla natura epistemologica del “Multilivello”

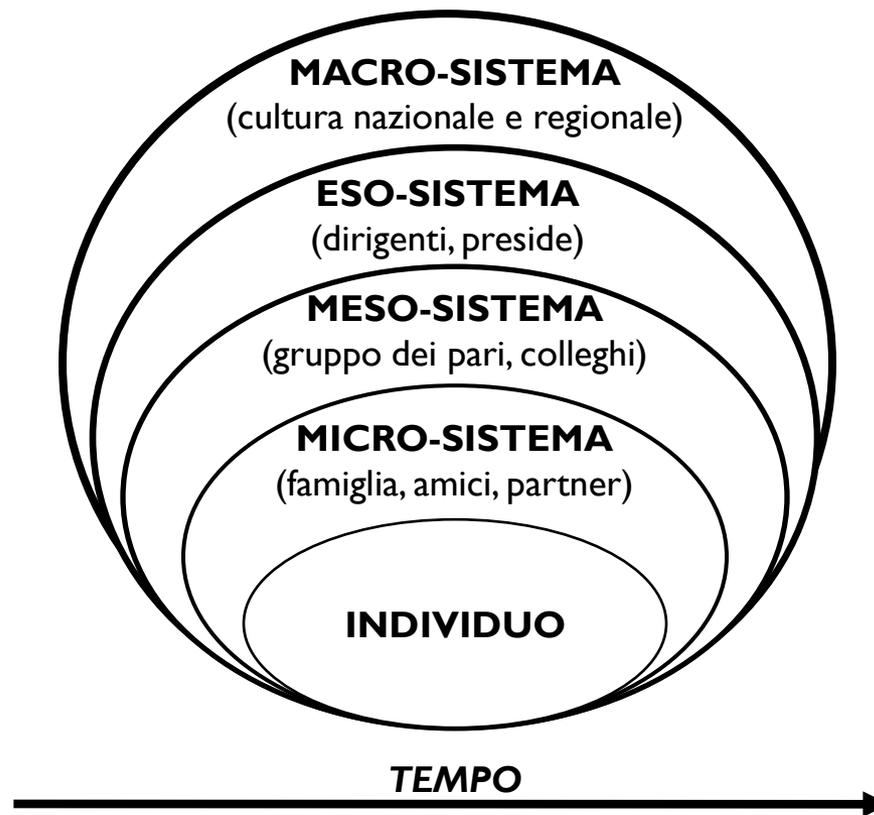
Alcuni aspetti relativi a tali strutture di dati pongono al ricercatore alcune considerazioni sulla “*nestedness*” dei dati e sulla relazione tra i livelli d’analisi.

In effetti:

- 1) Il semplice fatto che i dati siano “nested” in una unità d’analisi superiore non garantisce il fatto che la struttura dei dati sia multilivello;**
- 2) La natura del costrutto gioca un ruolo fondamentale (es., non si può considerare una variabile prettamente individuale come di livello sovraordinato);**
- 3) La relazione tra i livelli d’analisi non sempre è così chiara né tracciabile.**

Il modello dei sistemi sociali

Rispetto a questo, Bronfenbrenner (1975, 1977) ha sviluppato un modello della cosiddetta *ecologia sociale*, prevedendo una struttura dei livelli e delle relazioni di influenza da parte di fonti “esterne” sui processi di sviluppo individuale (vedi Little, Bovaird, & Card, 2007).



La “Shared Reality”

Tuttavia, non è sufficiente l'appartenenza a un determinato contesto sociale per “essere parte” di quel determinato contesto. A questo proposito prima Hardin e Higgins (1996) e successivamente Echterhoff, Higgins e Levine (2009) delineano alcune **caratteristiche alla base della condivisione della realtà psicologica e sociale**:

- 1) **La realtà viene condivisa** quando esiste quando non vengono **condivise alcune caratteristiche “covert”** (o “inner states”, es. il fastidio alla pressione del capo);
- 2) Tale condivisione deve avere un **target specifico di riferimento**;
- 3) **La condivisione** di cui al punto 1) **deve essere motivata** (i.e., avere una ragione identificabile);
- 4) La condivisione degli *inner states* deve **prevedere una connessione tra gli attori** del sistema sociale (es. la sua comunicazione).

I concetti di aggregazione e disaggregazione

Nel contesto dell'analisi multilivello, il concetto di **aggregazione** fa riferimento alla combinazione dei dati al livello gerarchico direttamente inferiore al fine di raggrupparle in un unico indice di livello superiore (es., la media valori di produttività individuale come variabile di livello di gruppo lavorativo) → Una variabile “WITHIN” viene considerata come “BETWEEN”

Il concetto di **disaggregazione** fa riferimento alla considerazione di alcuni dati propri livello gerarchico direttamente superiore al fine di utilizzare quell'informazione al livello gerarchico inferiore (es., risposta individuale a un item il cui target è il clima organizzativo) → Una variabile “BETWEEN” viene considerata come “WITHIN”

Variabili & Relazioni Multilivello

LIVELLO 3

MACRO LEVEL

Context
Composition
Structure
Resources



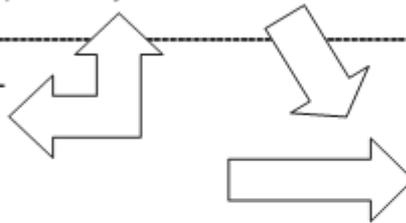
Organizational
Productivity

What contextual, structural, compositional, and resource variables affect organizational productivity?

LIVELLO 2

MACRO LEVEL

Context
Composition
Process



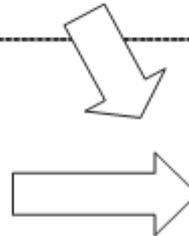
Departmental
Productivity

How do structural characteristics, compositional variables, and teamwork affect departmental productivity?

LIVELLO 1

MICRO LEVEL

Demographics
Attitudes
Previous Experiences



Individual
Productivity

How do background factors, attitudes, and previous experiences affect an employee's productivity?

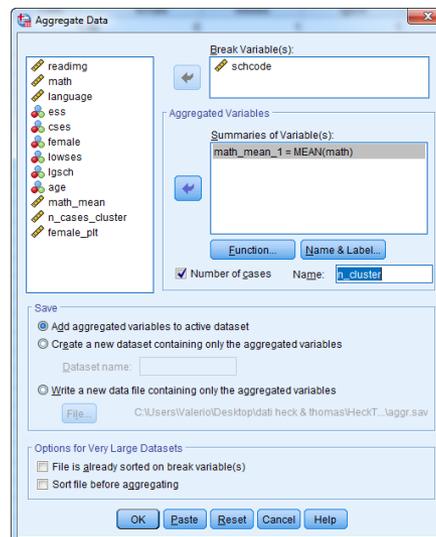
Adattato da Heck & Thomas, 2015, p. 4

Come strutturare i dati multilivello

ID LIV_1	ID LIV_3	ID LIV_2	Variabili L1			Variabile L2	Variabile L3
id_soggetto	id_scuola	id_classe	l1_a	l1_b	l1_c	l2_d	l3_e
1	1	100	682	714	673	-2,82	34
2	1	100	644	661	670	-2,82	34
3	1	100	651	670	648	-2,82	34
4	1	100	710	786	677	-2,82	34
5	1	100	673	719	698	-2,82	34
6	1	107	593	598	596	3,19	34
7	1	107	660	660	673	3,19	34
8	1	107	640	622	613	3,19	34
9	1	107	646	647	618	3,19	34
10	1	107	634	696	645	3,19	34
11	1	108	627	607	618	-5,43	34
12	1	108	620	635	648	-5,43	34
13	1	108	656	673	666	-5,43	34
14	1	108	606	614	613	-5,43	34
15	1	108	582	616	596	-5,43	34
16	2	109	649	676	663	3,94	21
17	2	109	646	688	681	3,94	21
18	2	109	662	685	673	3,94	21
19	2	109	647	665	670	3,94	21
20	2	109	715	690	681	3,94	21
21	2	111	610	644	629	3,29	21
22	2	111	640	635	663	3,29	21
23	2	111	685	652	654	3,29	21
24	2	111	660	630	643	3,29	21
25	2	111	634	665	640	3,29	21

AGGREGARE A LIVELLO 2 (SCHOOL) I PUNTEGGI NELLA VARIABILE MATH.

MENU “DATA” → FUNZIONE “AGGREGATE”
(molte possibilità di funzioni di aggregazione, come media, mediana, percentuali, ecc.)



AGGREGATE

/OUTFILE=* MODE=ADDVARIABLES *aggiunge la nuova variabile
*direttamente nel dataset in utilizzo

/BREAK=schcode *variabile “grouping” del livello

/math_mean=MEAN(math) *math_mean = nome della nuova variabile
*MEAN(math) = Funzione di
*aggregazione(variabile da “aggregare”

/n_cases_cluster=N. *restituisce una variabile che riflette la
*numerosità degli elementi “nested

**CREARE UNA VARIABILE DI LIVELLO 2 (SCHOOL)
CHE RIFLETTA LA PERCENTUALE DEI MASCHI
IN CIASCUNA SCUOLA, CHIAMANDOLA
LEV2_MALES_PERC**

La variabile di clustering è *schcode*

La variabile da aggregare è *female* (0 = maschio, 1 = femmina).

```
AGGREGATE  
  /OUTFILE=* MODE=ADDVARIABLES  
  /BREAK=schcode  
  /LEV2_MALES_PERC=PLT(female 1).
```

GIUSTIFICARE L'AGGREGAZIONE

Dal Livello 1 al Livello 2

Spesso, soprattutto nel contesto della ricerca psicologica e sociale, si hanno **ragioni teoricamente fondate per considerare alcune variabili misurate a livello individuale come degli indicatori di livello gerarchicamente superiore** (Rousseau, 1985);

Tendenzialmente, scelte del genere non vanno giustificate solo teoricamente ma **devono essere anche empiricamente tenibili** (Hox, 2010);

Di conseguenza, generalmente, vengono presentati alcuni indici che “quantificano” la possibilità di aggregare una variabile indipendente utilizzandola a un livello gerarchico superiore o per valutare quanta variabilità di una variabile dipendente è locata su altri livelli.

ICC & Deff

Esistono moltissimi indici in tal senso (si veda Bliese, 2000; LeBreton & Senter, 2007; Biemann, Cole, & Voelpel, 2012) che **valutano in realtà differenti aspetti della variabilità delle variabili attraverso i livelli**;

Quelli maggiormente usati sono il coefficiente di correlazione intraclassa (ICC o ICC(1)) e il Design Effect (DEff, Muthén & Satorra, 1995).

ICC o ICC(1)

$$\rho = \sigma^2_b / (\sigma^2_b + \sigma^2_w)$$

Dove:

σ^2_b = *varianza a livello di gruppo o between*

σ^2_w = *varianza a livello individuale o within*

Deff

$$\text{Deff} = [1 + (\gamma - 1 \times \rho)]$$

Dove:

γ = *numerosità media del cluster*;

ρ = *valore calcolato dell'ICC*

ICC & Deff

Mentre l'**ICC** riflette la proporzione di variabilità spiegata dalla **group membership** (Bryk & Raudenbush, 1982), il **DEff** quantifica sostanzialmente la dipendenza tra i soggetti, tecnicamente lo scarto di “casualità” (poiché essi sono “nested”) da un ipotetico campionamento casuale semplice;

Tendenzialmente, ai fini dell'aggregazione, vengono considerati accettabili valori del'ICC $>.10$ e del DEff > 2 , ma **questi valori possono non assumere significati profondamente diversi a seconda del :** 1) contesto di ricerca, 2) costrutti utilizzati, 3) rete nomologica dei modelli; 4) utilizzo o meno di variabili latenti;

Se, ad esempio, l'ICC = .01, significa nel caso di una VI che è difficile giustificarne l'aggregazione, nel caso di una VD che al livello superiore la varianza “spiegabile” da VI di quel livello è solo l'1% della variabilità totale dell'outcome.

CALCOLARE ICC e Deff per la variabile MATH

TITLE: EG2

DATA: FILE IS ch1example2.dat;
Format is 1f8.0,3f4.0,2f4.2,4f4.0;

VARIABLE: Names are schcode read math lang ess cses female
lowses lgsch age;

usevariables are math ;

CLUSTER IS schcode;

ANALYSIS: TYPE = Twolevel basic;

EG2 – Calcolo dell'ICC & del Deff – e2.inp

CALCOLARE ICC e DEff per la variabile *math*

SUMMARY OF DATA

```
Number of clusters                24

Size (s)   Cluster ID with Size s
   5       100 107 108 109 111 114 124 130 132 134 136 140 141
           143 145 150 151 153 155 250 251 274 375 377

Average cluster size              5.000

Estimated Intraclass Correlations for the Y Variables

Variable   Intraclass
           Correlation
MATH       0.359
```

ESTIMATED SAMPLE STATISTICS FOR WITHIN

```
Means
MATH
 1      0.000

Covariances
MATH
MATH    1257.009

Correlations
MATH
MATH    1.000
```

ESTIMATED SAMPLE STATISTICS FOR BETWEEN

```
Means
MATH
 1      667.192

Covariances
MATH
MATH    704.723

Correlations
MATH
MATH    1.000
```

$$\rho = \sigma_b^2 / (\sigma_b^2 + \sigma_w^2) = 704.723 / (704.723 + 1257.009) = .359$$

$$DEff = [1 + (\gamma - 1 \times \rho)] = [1 + (5 - 1 \times .359)] = 5.641$$

ES2 – Calcolo dell'ICC & del Deff

**PARTENDO DALLA SINTASSI EG2.inp,
CALCOLARE l'ICC per la variabile *age*
E IL DEff per la variabile *read***

SUMMARY OF DATA

```
Number of clusters                24

Size (s)   Cluster ID with Size s
5          100 107 108 109 111 114 124 130 132 134 136 140 141
          143 145 150 151 153 155 250 251 274 375 377

Average cluster size              5.000

Estimated Intraclass Correlations for the Y Variables

Variable   Intraclass
           Correlation
AGE        0.041
```

$$\text{DEff} = [1 + (\gamma - 1 \times \rho)] = [1 + (5 - 1 \times .142)] = 5.858$$

LA REGRESSIONE MULTIVELLO

La sequenza della regressione multilivello

Secondo Heck & Thomas (2015), i modelli da testare sequenzialmente nell'ambito della regressione multilivello (e, come vedremo più avanti, nella path analysis multilivello) sono:

- 1) Partizione della varianza dell'outcome** (o degli outcome se il modello è multivariato) **attraverso i livelli** (*Model 1* o *Modello nullo*);
- 2) Aggiungere predittori al livello di analisi più basso** (*Model 2*);
- 3) Aggiungere predittori al livello superiore** (*Model 3*);
- 4) Esaminare la variabilità delle slopes al livello superiore** (*Model 4*);
- 5) Aggiungere predittori al livello superiore per spiegare la variabilità nelle slopes** (*Model 5*).

Stima dei parametri & numerosità dei livelli

Mplus utilizza di default MLR, robusta a non esagerate violazioni della normalità multivariata, in congiunzione con l'approccio FIML alla gestione dei dati mancanti. **Questo metodo di stima dei parametri è ideale e preferibile nel caso di gruppi dalla numerosità anche fortemente sbilanciata;**

Quanti soggetti e quanti gruppi? Nonostante esistano alcune raccomandazioni, non esistono dei veri e propri cut-off di facile utilizzo per risolvere questo problema. Come sottolineato da Raudenbush e Bryk (2002), il fattore che maggiormente può creare dei problemi è la scarsità (ad esempio, se abbiamo 2 livelli) della N a livello 2, ed è fortemente dipendente dal tipo di modello da testare (Hox & Maas, 2001).

Esercitazioni “Guidate” (EG):

13189 impiegati *nested* 165 unità (Heck & Thomas, 2015, cap. 3)

Esercitazioni “Individuali” (ES):

503 studenti *nested* 34 classi (Geiser, 2013, cap. 5).

Equazioni del MODEL I (o Modello Nullo)

Questo modello ci permette di decomporre la varianza tra i livelli (WITHIN & BETWEEN), è matematicamente equivalente a una one-way ANOVA.

L'outcome è *morale*, misurata su una scala che va da 0 a 40.

$$\text{LIVELLO 1 } Y_{ij} = \beta_{0j} + \varepsilon_{ij}$$

$$\text{LIVELLO 2 } \beta_{0j} = \gamma_{00} + u_{0j}$$

Dove:

Y_{ij} = Punteggio dell'individuo in *morale* i nel gruppo j ;

B_{0j} = Media del punteggio del gruppo j ;

e_{ij} = Deviazione del punteggio del soggetto i dalla media del gruppo j ;

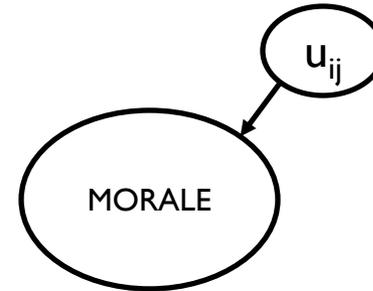
γ_{00} = Media dell'outcome tra tutti i gruppi;

u_{0j} = Deviazione del gruppo j dalla media dei gruppi.

Combinando le due equazioni, si ha che:

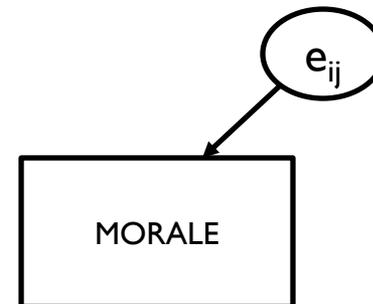
$$Y_{ij} = \gamma_{00} + u_{0j} + \varepsilon_{ij}$$

Rappresentazione del MODEL I (o Modello Nullo)



BETWEEN

WITHIN



EG3 - MODEL I - [ml.inp](#)

```
TITLE:          Model I:Two-level (null) regression model;
DATA:          FILE IS ch3new.dat;
                  Format is 5f8.0,3f8.2;

VARIABLE:     Names are deptid morale satpay female white pctbelow lev1 wt
                  lev2wt;
                  Usevariables are deptid morale;
                  Cluster is deptid;
                  Between = ;
                  Within = ;

ANALYSIS:     TYPE= Twolevel;
                  Estimator = MLR;

MODEL:
                  %Between%
                  morale;
                  %Within%
                  morale;

OUTPUT:       Sampstat Tech I;
```

EG3 - MODEL I- ml.out

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
within Level				
Variances MORALE	33.302	0.595	55.995	0.000
Between Level				
Means MORALE	26.428	0.189	139.945	0.000
Variances MORALE	5.363	0.608	8.817	0.000

$$Y_{ij} = \gamma_{00} + u_{0j} + \varepsilon_{ij}$$

La varianza è significativa sia a livello WITHIN che livello BETWEEN.

TECHNICAL 1 OUTPUT

PARAMETER SPECIFICATION FOR WITHIN

```
      NU
      MORALE
      _____
1      0
```

```
      THETA
      MORALE
      _____
MORALE 1
```

PARAMETER SPECIFICATION FOR BETWEEN

```
      NU
      MORALE
      _____
1      2
```

```
      THETA
      MORALE
      _____
MORALE 3
```

EG3 - MODEL I - Attendibilità del singolo gruppo

$$\lambda = \frac{\sigma_B^2}{\left[\sigma_B^2 + (\sigma_W^2 / n_j) \right]}$$

dove n_j rappresenta la numerosità de gruppo j .

Tale indice rappresenta l'attendibilità di uno specifico gruppo rispetto alla media tra i gruppi, ovvero quantifica l'attendibilità di una certa media rispetto a quella degli altri gruppi.

In un gruppo che ha 14 soggetti:

$$\lambda = 5.363 / [(5.363 + (33.302 / 14))] = 0.693$$

In un gruppo che ha 202 soggetti:

$$\lambda = 5.363 / [(5.363 + (33.302 / 202))] = 0.970$$

ES3 - MODEL I - [ml.inp](#)

Partendo dalla sintassi `ML.inp` nella cartella “Esercitazioni Guidate”, costruire un modello multilivello nullo (*Model I*) e valutare se la variabilità della variabile dipendente (*math*) è significativamente diversa da 0 sia a livello **WITHIN** che **BETWEEN**

```
MODEL RESULTS
```

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Within Level				
Variances				
MATH	26.480	2.450	10.809	0.000
Between Level				
Means				
MATH	11.934	0.808	14.775	0.000
Variances				
MATH	20.855	6.018	3.465	0.001

Equazioni del MODEL 2 (o RI Model)

Questo modello, detto anche *unconditional model* o *random-intercept (RI) model*, consente di introdurre dei predittori a livello 1 per spiegare la variabilità a livello WITHIN.

SATPAY = soddisfazione per la retribuzione, scala 0-16;

FEMALE = 0 maschi, 1 femmine;

WHITE = 0 altri, 1 bianchi.

$$\text{LIVELLO 1} \quad Y_{ij} = \beta_{0j} + \beta_{1j} \text{satpay}_{ij} + \beta_{2j} \text{female}_{ij} + \beta_{3j} \text{white}_{ij} + \varepsilon_{ij}$$

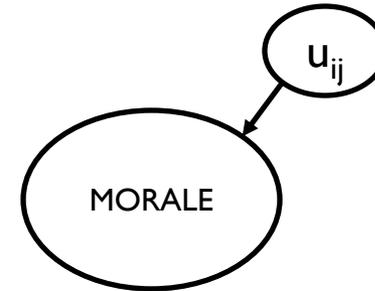
$$\text{LIVELLO 2} \quad \begin{aligned} \beta_{0j} &= \gamma_{00} + u_{0j} \\ \beta_{1j} - \beta_{3j} &= \gamma_{10} - \gamma_{30} \end{aligned}$$

Dove:

$\beta_{1j} \text{satpay}_{ij} - \beta_{2j} \text{female}_{ij}$ = Coefficiente di regressione dell'individuo i nel gruppo j ;

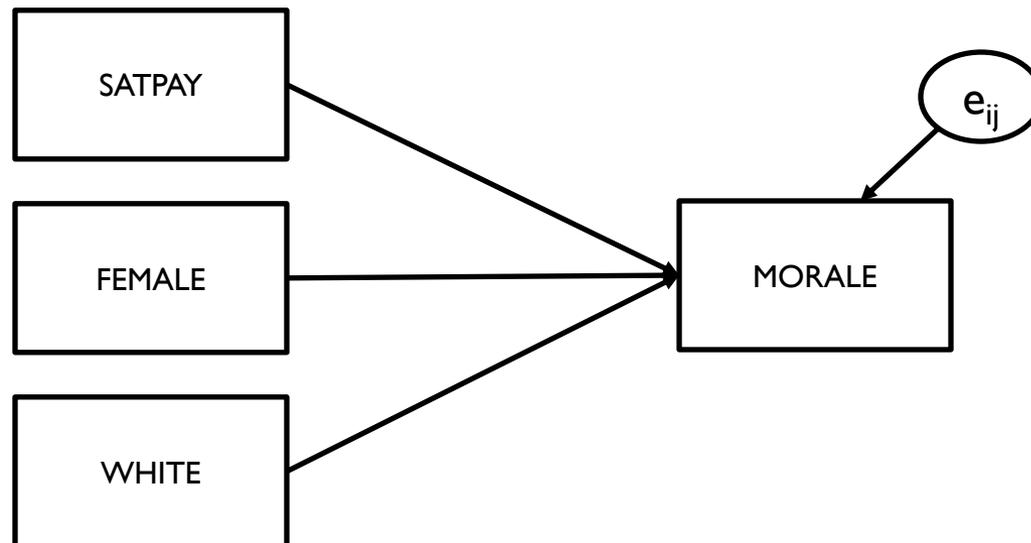
$\gamma_{00} - \gamma_{13}$ = Intercetta relativa al predittore di riferimento.

Rappresentazione del del MODEL 2 (o RI Model)



BETWEEN

WITHIN



EG4 - MODEL 2 - [m2.inp](#)

TITLE: Model 2: Level-1 random intercept model;

DATA: FILE IS ch3new.dat;
Format is 5f8.0,3f8.2;

VARIABLE: Names are deptid morale satpay female white pctbelow
lev1wt lev2wt ;
Usevariables are deptid morale satpay female white;
Cluster is deptid;
Between = ;
Within = satpay female white;
! specifica che le variabili non hanno variabilità al livello 2

Define: **Center satpay female white (grand);**
! i predittori vengono centrati alla media generale

ANALYSIS: **Type=Twolevel;**

Model: %Between%
morale;
%Within%
morale on satpay female white;

OUTPUT: SAMPSTAT TECHI;

EG4 - MODEL 2 – Centraggio dei predittori

Nell'analisi multilivello i predittori vanno centrati, siano essi dei predittori di livello 1 o livello 2. Centrare una variabile significa sottrarre al punteggio di un soggetto la media generale (GRANDMEAN CENTERING) o la media del gruppo di appartenenza (GROUPMEAN CENTERING).

GRANDMEAN CENTERING = I predittori vengono centrati sottraendo la media del campione dal punteggio. In questo modo, B_{0j} (intercetta) non subisce eventuali variazioni rilevanti rispetto a grandi variabilità interna al gruppo nei predittori e può considerarsi la media del gruppo j “adjusted”. **Si utilizza generalmente per i predittori di livello 1;**

GROUPMEAN CENTERING = I predittori vengono centrati sottraendo la media del gruppo dal punteggio. Questa strategia enfatizza eventuali differenze tra i gruppi, che si riflette in una minore variabilità a livello individuale. **Si utilizza generalmente per i predittori di livello 2 o sovraordinato.**

EG4 - MODEL 2- ml.out

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Within Level				
MORALE ON				
SATPAY	1.201	0.014	85.725	0.000
FEMALE	0.001	0.063	0.011	0.991
WHITE	0.916	0.082	11.181	0.000
Residual Variances				
MORALE	17.544	0.288	60.852	0.000
Between Level				
Means				
MORALE	26.430	0.114	232.784	0.000
Variances				
MORALE	1.851	0.231	8.029	0.000

LIVELLO 1

BETA	MORALE	SATPAY	FEMALE	WHITE
MORALE	0	1	2	3
SATPAY	0	0	0	0
FEMALE	0	0	0	0
WHITE	0	0	0	0

PSI	MORALE	SATPAY	FEMALE	WHITE
MORALE	4			
SATPAY	0	0		
FEMALE	0	0	0	
WHITE	0	0	0	0

LIVELLO 2

ALPHA	MORALE	SATPAY	FEMALE	WHITE
MORALE	5	0	0	0

PSI	MORALE	SATPAY	FEMALE	WHITE
MORALE	6			
SATPAY	0	0		
FEMALE	0	0	0	
WHITE	0	0	0	0

LIVELLO 1
$$Y_{ij} = \beta_{0j} + \beta_{1j} \text{satpay}_{ij} + \beta_{2j} \text{female}_{ij} + \beta_{3j} \text{white}_{ij} + \varepsilon_{ij}$$

$$\rho_{ri} = 1.851 / (1.851 + 17.544) = .095$$

LIVELLO 2
$$\beta_{0j} = \gamma_{00} + u_{0j}$$

riduzione nel $\rho = [(\rho_{\text{iniziale}} - \rho_{ri}) / \rho_{\text{iniziale}}] * 100 = 31.65\%$

$$\beta_{1j} - \beta_{3j} = \gamma_{10} - \gamma_{30}$$

La variabilità del livello 2 è diminuita del 31.65% (inizialmente l'ICC era .139)

ES4 - MODEL 2 - [m2.inp](#)

Partendo dalla sintassi `ML.inp`, costruire un RI model con `kft` come unico predittore, centrandolo alla media del campione. Quanto, in proporzione, diminuisce la variabilità di `math` (ICC originale .44) al livello 2 dopo aver introdotto il predittore?

MODEL RESULTS				
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Within Level				
MATH ON				
KFT	0.234	0.012	20.042	0.000
Residual Variances				
MATH	14.883	1.052	14.141	0.000
Between Level				
Means				
MATH	12.310	0.437	28.142	0.000
Variances				
MATH	5.403	1.612	3.352	0.001

$$\rho_{ri} = 5.403 / (14.883 + 5.403) = .266$$

$$\text{riduzione nel } \rho = [(.44 - .266) / .44] * 100 = 39.54\%$$

Equazioni del MODEL 3 (o IAO Model)

Questo modello, detto anche intercept-as-outcome(s) model, consente di introdurre dei predittori al livello gerarchico superiore con l'obiettivo di spiegare la variabilità dell'outcome locata a quel livello.

PCTBELOW = percentuale di dipendenti con uno stipendio sotto la media del gruppo di riferimento.

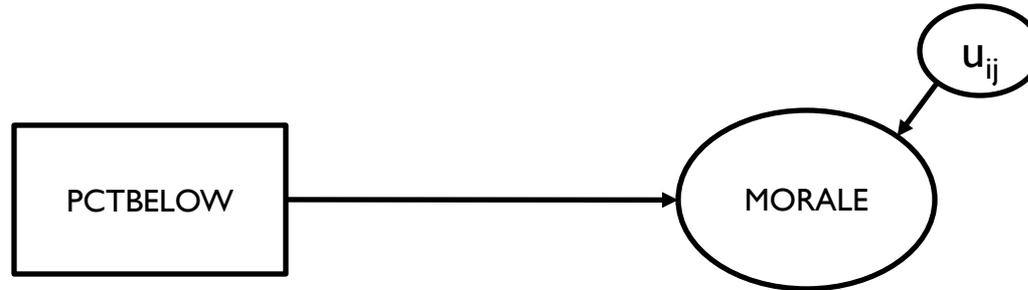
$$\text{LIVELLO 1} \quad Y_{ij} = \beta_{0j} + \beta_{1j} \text{satpay}_{ij} + \beta_{2j} \text{female}_{ij} + \beta_{3j} \text{white}_{ij} + \varepsilon_{ij}$$

$$\text{LIVELLO 2} \quad \beta_{0j} = \gamma_{00} + \gamma_{01} \text{pctbelow}_j + u_{0j}$$
$$\beta_{1j} - \beta_{3j} = \gamma_{10} - \gamma_{30}$$

Dove:

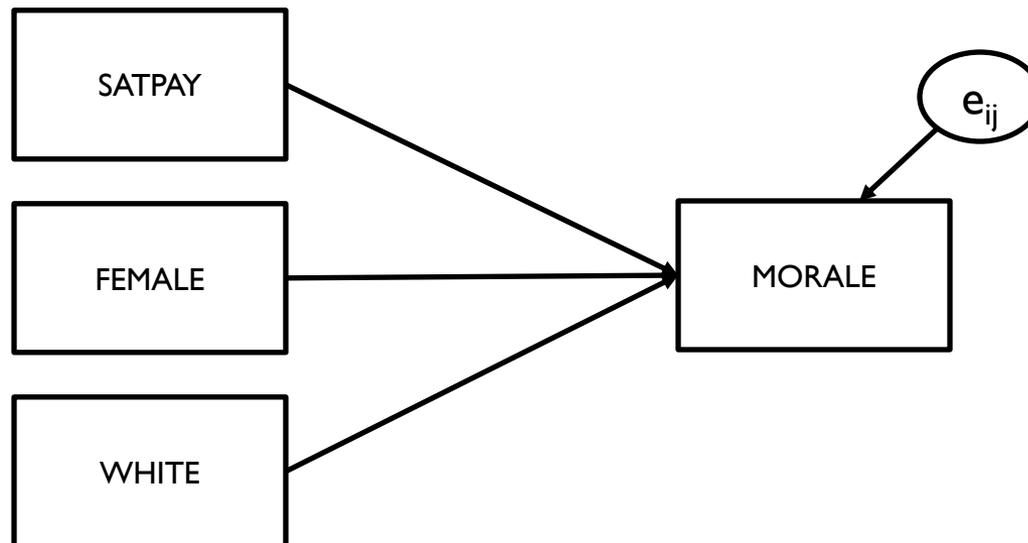
γ_{0j} = Coefficiente di regressione del gruppo j

Rappresentazione del del MODEL 3



BETWEEN

WITHIN



EG5 - MODEL 3 – [m3.inp](#)

TITLE: Model 3

DATA: FILE IS ch3new.dat;
Format is 5f8.0,3f8.2;

VARIABLE: Names are deptid morale satpay female white pctbelow
lev1wt lev2wt ;
Usevariables are deptid morale satpay female white;
Cluster is deptid;
Between = pctbelow;
!la variabile non ha variabilità a livello individuale
Within = satpay female white;

Define: **Center satpay female white pctbelow (grandmean);**
!pctbelow era stato precedentemente centrato alla media del gruppo, nel
!dataset tutti i soggetti di un gruppo hanno lo stesso valore

ANALYSIS: Type= Twolevel;

Model: **%Between%**
morale on pctbelow;
%Within%
morale on female white;

OUTPUT: SAMPSTAT TECHI;

EG4 - MODEL 3 – m3.out

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Within Level				
MORALE ON				
FEMALE	0.003	0.064	0.040	0.968
WHITE	0.910	0.082	11.046	0.000
SATPAY	1.200	0.014	85.315	0.000
Residual Variances				
MORALE	17.544	0.288	60.840	0.000
Between Level				
MORALE ON				
PCTBELOW	-0.027	0.007	-3.677	0.000
Intercepts				
MORALE	26.350	0.110	239.390	0.000
Residual Variances				
MORALE	1.675	0.215	7.779	0.000

$$\rho_{iao} = 1.675 / (17.544 + 1.675) = .087$$

riduzione nel $\rho = [(.139 - .087) / .139] * 100 = 37.54\%$

La variabilità del livello 2 è diminuita del 37.54% (l'introduzione di questo predittore aggiunge una quota di circa il 6% di varianza spiegata rispetto al modello precedente).

LIVELLO 1
$$Y_{ij} = \beta_{0j} + \beta_{1j} \text{satpay}_{ij} + \beta_{2j} \text{female}_{ij} + \beta_{3j} \text{white}_{ij} + \varepsilon_{ij}$$

LIVELLO 2
$$\beta_{0j} = \gamma_{00} + \gamma_{01} \text{pctbelow}_j + u_{0j}$$

$$\beta_{1j} - \beta_{3j} = \gamma_{10} - \gamma_{30}$$

EG4 - MODEL 3 – m3.out – TECH1

WITHIN

PARAMETER SPECIFICATION FOR WITHIN

BETA					
	MORALE	SATPAY	FEMALE	WHITE	PCTBELOW
MORALE	0	1	2	3	0
SATPAY	0	0	0	0	0
FEMALE	0	0	0	0	0
WHITE	0	0	0	0	0
PCTBELOW	0	0	0	0	0

PSI					
	MORALE	SATPAY	FEMALE	WHITE	PCTBELOW
MORALE	4				
SATPAY	0	0			
FEMALE	0	0	0		
WHITE	0	0	0	0	
PCTBELOW	0	0	0	0	0

BETWEEN

PARAMETER SPECIFICATION FOR BETWEEN

ALPHA					
	MORALE	SATPAY	FEMALE	WHITE	PCTBELOW
1	5	0	0	0	0

BETA					
	MORALE	SATPAY	FEMALE	WHITE	PCTBELOW
MORALE	0	0	0	0	6
SATPAY	0	0	0	0	0
FEMALE	0	0	0	0	0
WHITE	0	0	0	0	0
PCTBELOW	0	0	0	0	0

PSI					
	MORALE	SATPAY	FEMALE	WHITE	PCTBELOW
MORALE	7				
SATPAY	0	0			
FEMALE	0	0	0		
WHITE	0	0	0	0	
PCTBELOW	0	0	0	0	0

ES4 - MODEL 2 – [m3.inp](#)

Partendo dalla sintassi M2.inp, costruire un model con *kft* come unici predittori (LIVELLO 1) e *stype* (dicotomica di LIVELLO 2, non va centrata).

Quanto si riduce la variabilità al livello 2 rispetto al modello “nullo”?

Quanta variabilità in più spiega questo modello rispetto a quello testato il precedenza ($\rho = .266$)?

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Within Level				
MATH ON				
KFT	0.222	0.011	19.450	0.000
Residual Variances				
MATH	14.864	1.054	14.104	0.000
Between Level				
MATH ON				
STYPE	4.328	1.026	4.218	0.000
Intercepts				
MATH	11.488	0.390	29.470	0.000
Residual Variances				
MATH	2.904	0.716	4.058	0.000

$$\rho_{iao} = 2.904 / (14.864 + 2.904) = .163$$

$$\text{riduzione nel } \rho = [(.44 - .163) / .44] * 100 = 66.95\%$$

Spiega in più circa il 23% della variabilità del livello 2 rispetto al modello precedente

Analisi Multilivello con *Mplus*

Seminario I – 9 Maggio 2016

Valerio Ghezzi

Dipartimento di Psicologia

Sapienza – Università di Roma