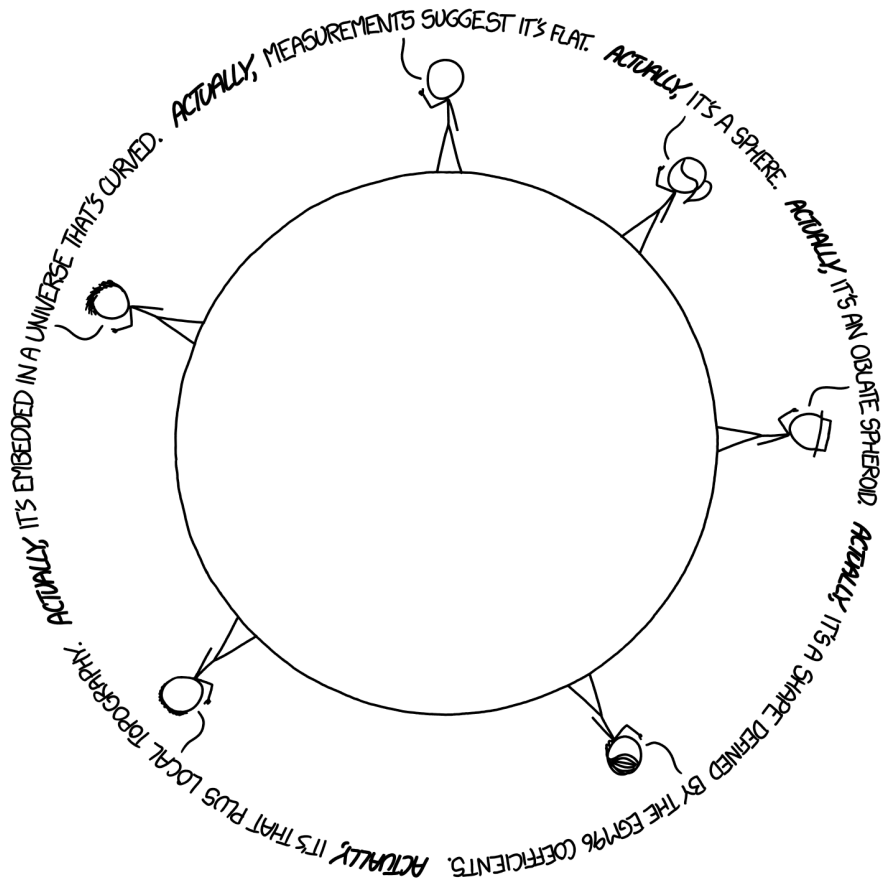


Corrado Caudek

***Appunti di Costruzione e
validazione di strumenti di
misura dell'efficacia
dell'intervento psicologico
in neuropsicologia –
B020881 (B213)***



Appunti di Costruzione e validazione di strumenti di
misura dell'efficacia dell'intervento psicologico in
neuropsicologia – AA 2021/2022





Indice

Elenco delle figure	vii
Elenco delle tabelle	ix
Prefazione	xi
I Il modello lineare	1
1 L'analisi di regressione	3
1.1 Regressione bivariata	3
1.1.1 Regressori centrati	5
1.1.2 Minimi quadrati	6
1.1.3 Relazione tra b e r	8
1.1.4 Attenuazione	8
1.1.5 Coefficiente di determinazione	10
1.1.6 Errore standard della regressione	11
1.2 Regressione multipla	11
1.2.1 Significato dei coefficienti parziali di regressione .	12
1.2.2 Relazioni causali	14
1.2.3 Errore di specificazione	15
1.2.4 Soppressione	18
1.2.5 Stepwise regression	19
2 Invarianza di misura	21
2.1 Indicatori continui	21
2.1.1 Intercette degli item	21
2.1.2 Terminologia	23
2.1.3 Un esempio concreto	23
2.2 Variabili a livello ordinale	26
2.2.1 Baseline model	28
2.2.2 Invarianza delle soglie	29
2.2.3 Threshold and loading invariance	31



Elenco delle figure



Elenco delle tabelle



Prefazione

La presente dispensa contiene il materiale delle lezioni dell'insegnamento di *Costruzione e validazione di strumenti di misura dell'efficacia dell'intervento psicologico in neuropsicologia* B020881 (B213) rivolto agli studenti del secondo anno del Corso di Laurea Magistrale in Psicologia Clinica e della Salute e Neuropsicologia (curriculum: assessment e intervento psicologici in neuropsicologia - E21), A.A. 2021-2022. L'insegnamento si propone di fornire agli studenti un'introduzione all'assessment psicologico, ovvero un insieme di conoscenze/competenze che si pongono all'intersezione tra psicometria, statistica e informatica.

Nello specifico, l'insegnamento si focalizzerà sull'analisi fattoriale confermativa (*confirmatory factor analysis*, CFA) e sull'analisi fattoriale esplorativa, (*explorative factor analysis*, EFA), cioè sugli strumenti che vengono usati durante il processo di sviluppo dei test psicometrici, ovvero che vengono usati per esaminare la struttura latente di una scala psicologica (ad esempio un questionario). In questo contesto, la CFA viene utilizzata per verificare il numero di dimensioni sottostanti gli indicatori (fattori) e l'intensità delle relazioni item-fattore (saturazioni fattoriali). La CFA consente anche di capire di come dovrebbe essere svolto lo scoring di un test. Quando la struttura latente è multifattoriale (cioè, a due o più fattori), il numero di fattori è indicativo del numero di sottoscale e di come esse dovrebbero essere codificate. La CFA è un importante strumento analitico anche per altri aspetti della valutazione psicometrica. Può essere utilizzata per stimare l'affidabilità di scala dei test psicometrici in modo da evitare i problemi della teoria classica dei test (ad es. alpha di Cronbach). Dati i recenti progressi nell'analisi dei dati categoriali, ora la CFA offre un quadro analitico comparabile a quello offerto dalla teoria di risposta agli item (IRT). In effetti, secondo [Brown \(2015\)](#), la CFA offre una maggiore flessibilità analitica rispetto al modello IRT tradizionale.

Un costrutto è un concetto teorico che può essere operazionalizzato nei termini di un fattore. In psicologia clinica, psichiatria e neuropsicologia,

ad esempio, i disturbi mentali sono costrutti manifestati da vari insiemi di sintomi che sono riportati dal paziente o osservati da altri. La CFA è uno strumento analitico indispensabile per la validazione dei costrutti psicologici. I risultati della CFA possono fornire prove convincenti della validità convergente e discriminante dei costrutti teorici. La validità convergente è indicata dall'evidenza che diversi indicatori di costrutti teoricamente simili o sovrapposti sono fortemente correlati. La validità discriminante è indicata dai risultati che mostrano che gli indicatori di costrutti teoricamente distinti sono altamente incorrelati. Un punto di forza fondamentale degli approcci CFA per la costruzione e la validazione di uno strumento psicometrico è che le risultanti stime di validità convergente e discriminante sono corrette per l'errore di misurazione. Pertanto, la CFA fornisce un quadro analitico migliore rispetto ai metodi tradizionali che non tengono conto dell'errore di misurazione (ad esempio, gli approcci ordinari ai minimi quadrati come la correlazione/regressione multipla, i quali presuppongono che le variabili nell'analisi siano prive di errori di misurazione).

Spesso, parte della covariazione delle misure osservate è dovuta a fonti diverse dai fattori latenti di interesse. Questa covariazione aggiuntiva spesso riflette la varianza del metodo utilizzato per la misurazione. Gli effetti del metodo possono verificarsi anche all'interno di un'unica modalità di valutazione. Ad esempio, effetti del metodo sono solitamente presenti nei questionari che contengono una combinazione di elementi formulati positivamente e negativamente. Sfortunatamente, l'EFA non è in grado di stimare gli effetti del metodo. In effetti, l'uso di EFA quando esistono effetti del metodo può produrre risultati fuorvianti, ovvero suggerire la presenza di fattori aggiuntivi che corrispondono invece ad artefatti della misurazione. Nella CFA, invece, gli effetti del metodo possono essere specificati come parte della teoria dell'errore del modello di misurazione.

Un altro punto di forza della CFA è la sua capacità di affrontare il problema della generalizzabilità del modello di misurazione tra gruppi di individui o nel tempo. La valutazione dell'invarianza della misura è un aspetto importante dello sviluppo del test. Se un test è destinato a essere somministrato in una popolazione eterogenea, si dovrebbe stabilire che le sue proprietà di misurazione sono equivalenti in sottogruppi della popolazione (es. sesso, razza). Si dice che un test è distorto quando alcuni dei suoi elementi non misurano il costrutto sottostante in modo comparabile tra gruppi di rispondenti. Il test fornisce una stima distorta se,

ad esempio, per un dato livello di vera intelligenza, gli uomini tendono a ottenere un punteggio di QI più alto rispetto alle donne. Il problema della generalizzabilità della validità del costrutto tra i gruppi può essere affrontato nella CFA esaminando gruppi multipli mediante modelli MI-MIC (indicatori multipli, cause multiple). Inoltre, è possibile chiedersi se il modello di misurazione sia equivalente tra i gruppi. Le soluzioni CFA a gruppi multipli vengono anche utilizzate per esaminare l'invarianza della misurazione longitudinale. Questo è un aspetto molto importante dell'analisi delle variabili latenti dei progetti di misure ripetute. In assenza di tale valutazione, non è possibile determinare se il cambiamento temporale in un costrutto sia dovuto a un vero cambiamento dei rispondenti o a cambiamenti nel modo di rispondere alla scala nel tempo. L'analisi a gruppi multipli può essere applicata a qualsiasi tipo di modello CFA. Ad esempio, queste procedure possono essere incorporate nell'analisi dei dati multitratto-multimetodo per esaminare la generalizzabilità della validità del costrutto tra gruppi.

In questo insegnamento la discussione delle tecniche della CFA sarà preceduta da un'introduzione relativa alla EFA e la teoria classica dei test. La EFA, infatti, può essere concepita il metodo che viene utilizzato nei primi passi dello sviluppo di una scala psicometria, mentre la teoria classica dei test rappresenta la cornice teorica di partenza, di cui la CFA e i modelli di equazioni strutturali costituiscono uno sviluppo.

L'insegnamento pone una grande enfasi non solo sulla comprensione dei concetti teorici necessari per la costruzione e la validazione di uno strumento di misura in psicologia, ma anche sulla capacità di applicare tali concetti in situazioni concrete. Di conseguenza, la discussione dei concetti sarà sempre accompagnata da applicazioni pratiche. Tali applicazioni richiedono l'uso di un software. In questo insegnamento useremo R ([R Core Team, 2021](#)) quale linguaggio di programmazione probabilistica e, tra gli altri, il pacchetto `lavaan` che consente di svolgere le analisi statistiche della CFA e della EFA ([Beaujean, 2014](#)). La teoria classica dei test verrà descritta con riferimento al classico testo di [Lord and Novick \(1968\)](#). Questa dispensa, inoltre, segue da vicino la trattazione della CFA fornita nei testi di [McDonald \(2013\)](#) e di [Brown \(2015\)](#).

Trattando di argomenti avanzati, questo insegnamento presuppone la conoscenza di base dei concetti fondamentali della teoria delle probabilità; presuppone inoltre il possesso delle conoscenze di base necessarie per procedere all'utilizzo di R. Informazioni su tali argomenti sono forniti

xiv

nella dispensa di Psicometria (A.A. 2021-2022).

Prefazione

Corrado Caudek
Marzo 2022

Parte I

Il modello lineare



1

L'analisi di regressione

Conoscere l'analisi di regressione aiuta a capire la teoria classica dei test, l'analisi fattoriale e i modelli di equazioni strutturali. Sebbene le tecniche dell'analisi di regressione analizzino solo le variabili osservate, i principi della regressione costituiscono la base delle tecniche più avanzate che includono anche le variabili latenti.

1.1 Regressione bivariata

Il modello di regressione bivariata descrive l'associazione tra il valore atteso di $Y \mid x_i$ e x nei termini di una relazione lineare:

$$\mathbb{E}(Y \mid x_i) = \alpha + \beta x_i,$$

dove i valori x_i sono considerati fissi per disegno. Nel modello “classico”, si assume che le distribuzioni $Y \mid x_i$ siano Normali con deviazione standard σ_ε .

Il significato dei coefficienti di regressione è semplice:

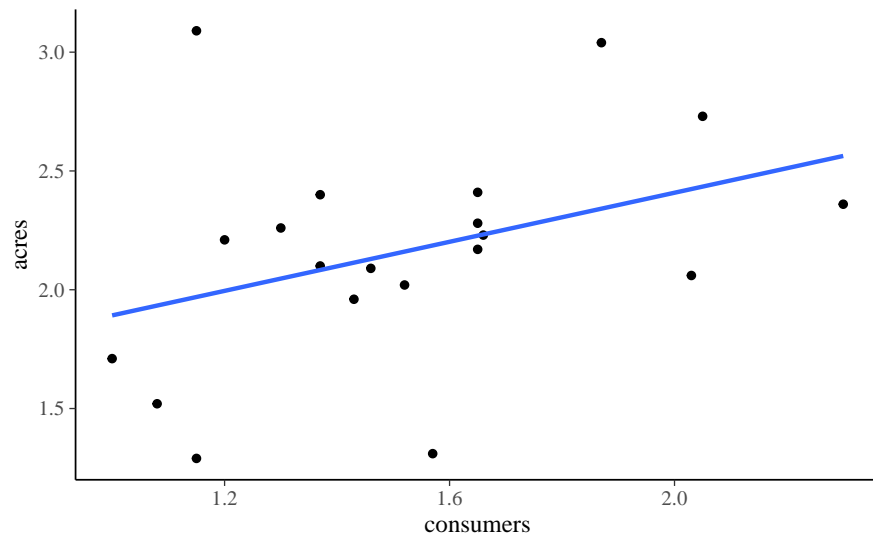
- α è il valore atteso di Y quando $X = 0$;
- β è l'incremento atteso nel valore atteso di Y quando X aumenta di un'unità.

Per fare un esempio, consideriamo i dati dell'antropologo Sahlins, il quale si è chiesto se esiste un'associazione tra l'ampiezza del clan (**consumers**) e l'area occupata da quel clan (**acres**) in una popolazione di cacciatori-raccoglitori. I dati sono i seguenti:

```
data(Sahlins)
head(Sahlins)
```

```
#>   consumers acres
#> 1      1.00  1.71
#> 2      1.08  1.52
#> 3      1.15  1.29
#> 4      1.15  3.09
#> 5      1.20  2.21
#> 6      1.30  2.26
```

```
Sahlins %>%
  ggplot(aes(x = consumers, y = acres)) +
  geom_point() +
  geom_smooth(method = lm, se = FALSE)
```



```
fm <- lm(acres ~ consumers, data = Sahlins)
fm$coef
#> (Intercept)   consumers
#>      1.3756      0.5163
```

Dalla figura notiamo che, se `consumers` aumenta di un'unità (da 1.2 a 2.2), allora la retta di regressione (ovvero, il valore atteso di Y) aumenta di circa 0.5 punti – esattamente, aumenta di 0.5163 punti, come indicato

dalla stima del coefficiente β . L'interpretazione del coefficiente α è più problematica, perché non ha senso pensare ad un clan di ampiezza 0. Per affrontare questo problema, centriamo il predittore.

1.1.1 Regressori centrati

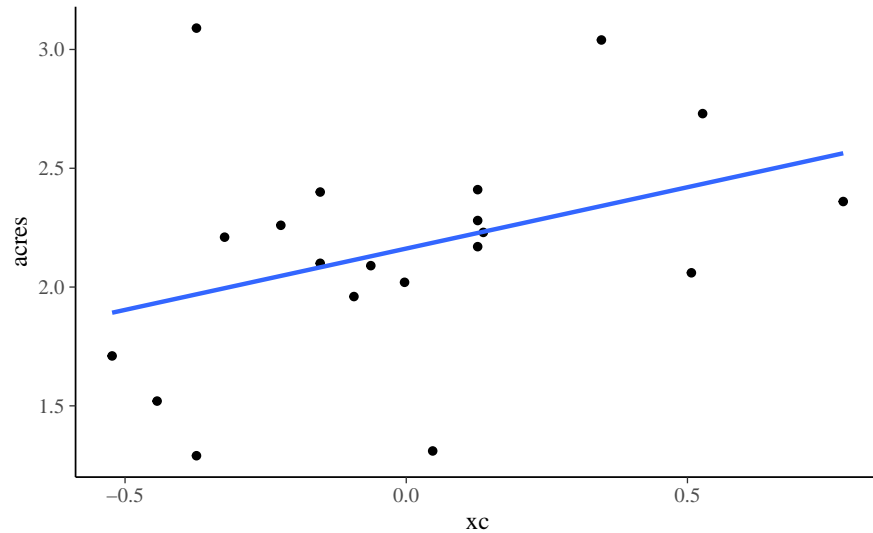
Esprimiamo la variabile `consumers` nei termini degli scarti dalla media:

```
Sahlins <- Sahlins %>%  
  mutate(  
    xc = consumers - mean(consumers)  
  )
```

Svolgiamo nuovamente l'analisi di regressione con il nuovo predittore:

```
fm1 <- lm(acres ~ xc, data = Sahlins)  
fm1$coef  
#> (Intercept)          xc  
#>      2.1620      0.5163
```

```
Sahlins %>%  
  ggplot(aes(x = xc, y = acres)) +  
  geom_point() +  
  geom_smooth(method = lm, se = FALSE)
```



La stima di β è rimasta invariata ma ora possiamo attribuire un significato alla stima di α : questo coefficiente indica il valore atteso della Y quando X assume il suo valore medio.

1.1.2 Minimi quadrati

La stima dei coefficienti del modello di regressione può essere effettuata in modi diversi: massima verosimiglianza o metodi bayesiani. Se ci limitiamo qui alla massima verosimiglianza possiamo semplificare il problema assumendo che le distribuzioni condizionate $Y | x$ siano Normali. In tali circostanze, la stima dei coefficienti del modello di regressione può essere trovata con il metodo dei minimi quadrati.

In pratica, questo significa trovare i coefficienti a e b che minimizzano

$$SS_{\text{res}} = \sum (y_i - \hat{y}_i)^2,$$

con $\hat{y}_i = a + bx_i$.

Per fornire un'idea di come questo viene fatto, usiamo una simulazione. Per semplicità, supponiamo di conoscere $a = 1.3756445$ e di volere stimare b .

```
x <- Sahlins$consumers
y <- Sahlins$acres
```

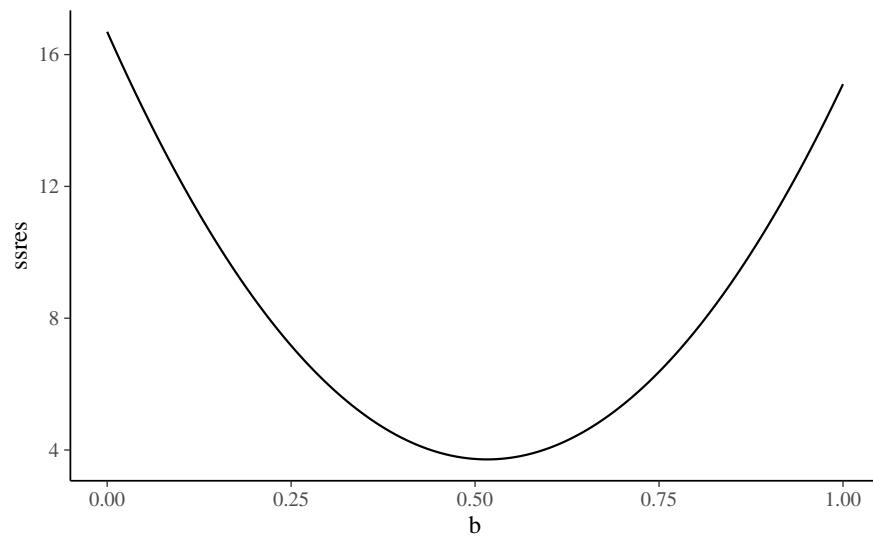
```
a <- 1.3756445

nrep <- 1e3
b <- seq(0, 1, length.out = nrep)

ssres <- rep(NA, nrep)
for (i in 1:nrep) {
  yhat <- a + b[i] * x
  ssres[i] <- sum((y - yhat)^2)
}
```

Un grafico di SS_{res} in funzione di b mostra che il valore b che minimizza SS_{res} corrisponde, appunto, a 0.5163.

```
tibble(b, ssres) %>%
  ggplot(aes(x = b, y = ssres)) +
  geom_line()
```



1.1.3 Relazione tra b e r

Un altro modo per interpretare b è quello di considerare la relazione tra la pendenza della retta di regressione e il coefficiente di correlazione:

$$b_X = r_{XY} \frac{S_X}{S_Y}$$

L'equazione precedente rende chiaro che, se i dati sono standardizzati, $b = r$.

Verifichiamo:

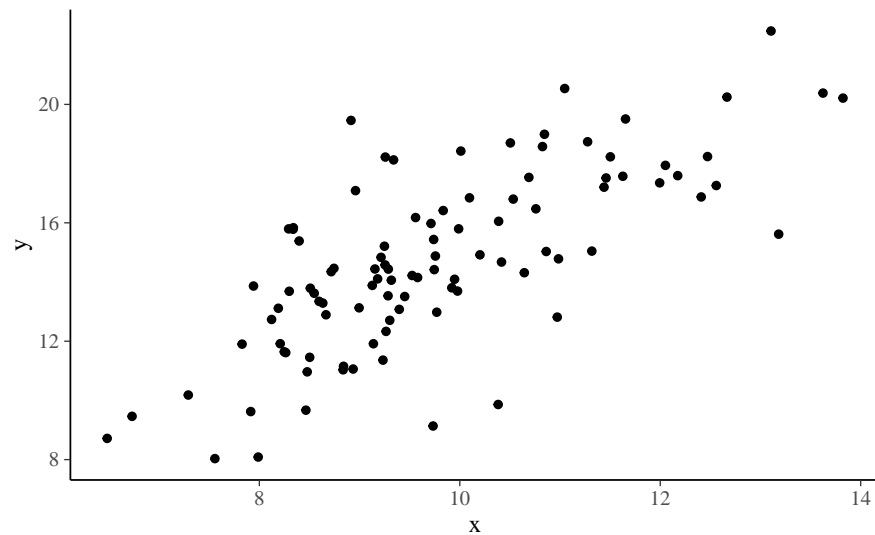
```
Sahlins %>%
  dplyr::select(acres, consumers) %>%
  cor()
#>           acres consumers
#> acres      1.0000    0.3757
#> consumers 0.3757    1.0000
```

```
fm2 <- lm(scale(acres) ~ scale(consumers), data = Sahlins)
fm2$coef
#>      (Intercept) scale(consumers)
#>      9.917e-17      3.757e-01
```

1.1.4 Attenuazione

Il fenomeno dell'attenuazione si verifica quando X viene misurato con una componente di errore. Esaminiamo la seguente simulazione.

```
set.seed(1234)
n <- 100
x <- rnorm(n, 10, 1.5)
y <- 1.5 * x + rnorm(n, 0, 2)
tibble(x, y) %>%
  ggplot(aes(x, y)) +
  geom_point()
```



```
sim_dat <- tibble(x, y)
fm <- lm(y ~ x, sim_dat)
fm$coef
#> (Intercept)          x
#>      0.4221      1.4652
```

Questi sono i coefficienti di regressione quando X è misurata senza errori.

```
sim_dat <- sim_dat %>%
  mutate(
    x1 = x + rnorm(n, 0, 2)
  )

fm1 <- lm(y ~ x1, sim_dat)
fm1$coef
#> (Intercept)          x1
#>      8.3872      0.6296
```

Aggiungendo una componente d'errore su X , la grandezza del coefficiente b diminuisce.

1.1.5 Coefficiente di determinazione

Tecnicamente, il coefficiente di determinazione è dato da:

$$R^2 = \frac{\sum(\hat{y} - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

Al denominatore abbiamo la *devianza totale*, ovvero una misura della dispersione di y_i rispetto alla media \bar{y} . Al numeratore abbiamo una misura della dispersione del valore atteso della Y rispetto alla sua media. Il rapporto, dunque, ci dice qual è la quota della variabilità totale di Y che può essere predetta in base al modello lineare.

Per i dati di Sahlins abbiamo:

```
mod <- lm(acres ~ consumers, data = Sahlins)
a <- mod$coef[1]
b <- mod$coef[2]
yhat <- a + b * Sahlins$consumers
ss_tot <- sum((Sahlins$acres - mean(Sahlins$acres))^2)
ss_reg <- sum((yhat - mean(Sahlins$acres))^2)
r2 <- ss_reg / ss_tot
r2
#> [1] 0.1411
```

Verifichiamo:

```
summary(mod)
#>
#> Call:
#> lm(formula = acres ~ consumers, data = Sahlins)
#>
#> Residuals:
#>      Min       1Q   Median       3Q      Max
#> -0.8763 -0.1873 -0.0211  0.2135  1.1206
#>
#> Coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)    1.376      0.468    2.94   0.0088 **
#> consumers      0.516      0.300    1.72   0.1026
```



```
#> ---
#> Signif. codes:
#> 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 0.454 on 18 degrees of freedom
#> Multiple R-squared: 0.141, Adjusted R-squared: 0.0934
#> F-statistic: 2.96 on 1 and 18 DF, p-value: 0.103
```

Da cui deriva che R^2 è uguale al quadrato del coefficiente di correlazione:

```
cor(Sahlins$acres, Sahlins$consumers)^2
#> [1] 0.1411
```

1.1.6 Errore standard della regressione

L'errore standard della regressione è una stima della dispersione di $y \mid x_i$ nella popolazione. Non è altro che la deviazione standard dei residui

$$e = y_i - \hat{y}_i$$

che, al denominatore, riporta $n - 2$. La ragione è che, per calcolare \hat{y} , vengono “perduti” due gradi di libertà – il calcolo di \hat{y} è basato sulla stima di due coefficienti: a e b .

```
e <- yhat - Sahlins$acres
(sum(e^2) / (length(Sahlins$acres) - 2)) %>%
  sqrt()
#> [1] 0.4543
```

Il valore trovato corrisponde a quello riportato nell'output di `lm()`.

1.2 Regressione multipla

Nella regressione multipla vengono utilizzati $k > 1$ predittori:

$$y_i = \alpha + \sum_{j=1}^k \beta_j x_i + \varepsilon_i.$$

L'interpretazione geometrica è simile a quella del modello bivariato. Nel caso di due predittori, il valore atteso della y può essere rappresentato da un piano; nel caso di $k > 2$ predittori, da un iper-piano. Nel caso di $k = 2$, tale piano è posto in uno spazio di dimensioni x_1, x_2 (che possiamo immaginare definire un piano orizzontale) e y (ortogonale a tale piano). La superficie piana che rappresenta $\mathbb{E}(y)$ è inclinata in maniera tale che l'angolo tra il piano e l'asse x_1 corrisponde a β_1 e l'angolo tra il piano e l'asse x_2 corrisponde a β_2 .

1.2.1 Significato dei coefficienti parziali di regressione

Ai coefficienti parziali del modello di regressione multipla possiamo assegnare la seguente interpretazione:

Il coefficiente parziale di regressione β_j rappresenta l'incremento atteso della y se x_j viene incrementata di un'unità, tenendo costante il valore delle altre variabili indipendenti.

Un modo per interpretare la locuzione “al netto dell'effetto delle altre variabili indipendenti” è quello di esaminare la relazione tra la y parzializzata e la x_j parzializzata. In questo contesto, parzializzare significa decomporre una variabile di due componenti: una componente che è linearmente predicibile da una o più altre variabili e una componente che è linearmente incorrelata con tali variabili “terze”.

Eseguiamo questa “depurazione” dell'effetto delle variabili “terze” sia sulla y sia su x_j . A questo punto possiamo esaminare la relazione bivariata che intercorre tra la componente della y linearmente indipendente dalle variabili “terze” e la componente della x_j linearmente indipendente dalle variabili “terze”. Il coefficiente di regressione bivariato così ottenuto è identico al coefficiente parziale di regressione nel modello di regressione multipla. Possiamo così ottenere un'interpretazione di β_j .

Esaminiamo un caso concreto.

```
d <- rio::import(
  here::here("data", "kidiq.dta")
)
```

```
glimpse(d)
#> Rows: 434
#> Columns: 5
#> $ kid_score <dbl> 65, 98, 85, 83, 115, 98, 69, 106, 1~
#> $ mom_hs    <dbl> 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1,~
#> $ mom_iq    <dbl> 121.12, 89.36, 115.44, 99.45, 92.75~
#> $ mom_work  <dbl> 4, 4, 4, 3, 4, 1, 4, 3, 1, 1, 1, 4,~
#> $ mom_age   <dbl> 27, 25, 27, 25, 27, 18, 20, 23, 24,~
```

```
fm <- lm(
  kid_score ~ mom_iq + mom_work + mom_age + mom_hs,
  data = d
)
fm$coef
#> (Intercept)      mom_iq      mom_work      mom_age
#>    20.8226     0.5621     0.1337     0.2199
#>      mom_hs
#>     5.5612
```

Eseguiamo la parzializzazione di y in funzione delle variabili `mom_work`, `mom_age` e `mom_hs`:

```
fm_y <- lm(kid_score ~ mom_work + mom_age + mom_hs, data = d)
```

Lo stesso per `mom_iq`:

```
fm_x <- lm(mom_iq ~ mom_work + mom_age + mom_hs, data = d)
```

Esaminiamo ora la regressione bivariata tra le componenti parzializzate della y e di x_j :

```
mod <- lm(fm_y$residuals ~ fm_x$residuals)
mod$coef
#> (Intercept) fm_x$residuals
#> -1.652e-15      5.621e-01
```

Si vede come il coefficiente di regressione bivariato risulta identico al corrispondente coefficiente parziale di regressione.

1.2.2 Relazioni causali

Un altro modo per interpretare i coefficienti parziali di regressione è nell'ambito dei quelli che vengono chiamati i *path diagrams*. I diagrammi di percorso, che tratteremo in seguito e qui solo anticipiamo, descrivono le relazioni “causali” tra variabili: le variabili a monte del diagramma di percorso indicano le “cause” esogene e le variabili a valle indicano gli effetti, ovvero le variabili endogene. I coefficienti di percorso rappresentati graficamente come frecce orientate corrispondono all'effetto *diretto* sulla variabile verso cui punta la freccia della variabile a monte della freccia. Tali coefficienti di percorso non sono altro che i coefficienti parziali di regressione del modello di regressione multipla. In questo contesto, indicano l'effetto atteso *diretto* sulla variabile endogena dell'incremento di un'unità della variabile esogena, lasciano immutate tutte le altre relazioni strutturali del modello.

Usiamo la funzione `sem()` del pacchetto `lavaan` per definire il modello rappresentato nel successivo diagramma di percorso:

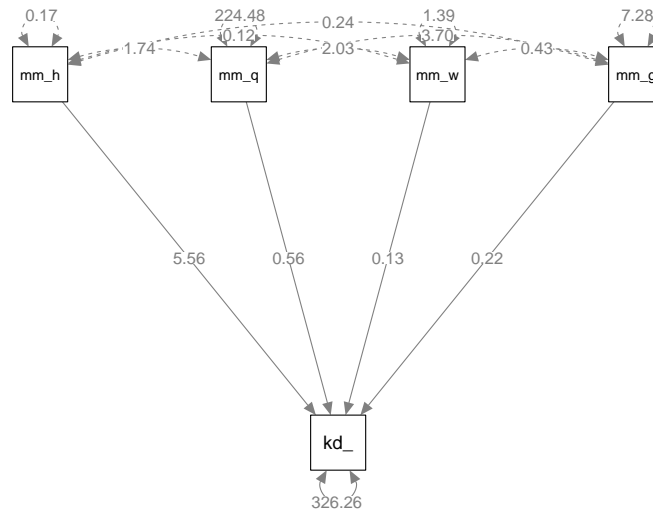
```
model <- "
  kid_score ~ mom_hs + mom_iq + mom_work + mom_age
"
```

Adattiamo il modello ai dati

```
fit <- sem(model, data = d)
```

Il diagramma di percorso si ottiene con le seguenti istruzioni:

```
semPaths(
  fit, "est",
  posCol = c("black"),
  edge.label.cex = 0.9,
  sizeMan = 7,
  what = "path"
)
```



Come indicato nel diagramma, l'effetto diretto di mm_q su kd_score è identico al corrispondente coefficiente parziale di regressione.

1.2.3 Errore di specificazione

Spiritualmente chiamato “heartbreak of L.O.V.E.” (Left-Out Variable Error; [Mauro \(1990\)](#)), l'errore di specificazione è una caratteristica fondamentale dei modelli di regressione che deve sempre essere tenuta a mente quando interpretiamo i risultati di questa analisi statistica. L'errore di specificazione si verifica quando escludiamo dal modello di regressione una variabile che

- è associata con altre variabili nel modello,
- ha un effetto diretto sulla y .

Come conseguenza dell'errore di specificazione, la direzione e il segno dei coefficienti parziali di regressione risultano sistematicamente distorti.

Consideriamo un esempio con dati simulati nei quali immaginiamo che la prestazione sia positivamente associata alla motivazione e negativamente associata all'ansia. Immaginiamo inoltre che vi sia una correlazione positiva tra ansia e motivazione. Ci chiediamo cosa succede al coefficiente parziale della variabile “motivazione” se la variabile “ansia” viene esclusa dal modello di regressione.

```
set.seed(123)
n <- 400
anxiety <- rnorm(n, 10, 1.5)
motivation <- 4.0 * anxiety + rnorm(n, 0, 3.5)
cor(anxiety, motivation)
#> [1] 0.8618
```

```
performance <- 0.5 * motivation - 5.0 * anxiety + rnorm(n, 0, 3)
```

```
sim_dat2 <- tibble(performance, motivation, anxiety)
fm1 <- lm(performance ~ motivation + anxiety, sim_dat2)
coef(fm1)
#> (Intercept) motivation anxiety
#> 1.3712 0.4954 -5.1052
```

```
fm2 <- lm(performance ~ motivation, sim_dat2)
summary(fm2)
#>
#> Call:
#> lm(formula = performance ~ motivation, data = sim_dat2)
#>
#> Residuals:
#> Min 1Q Median 3Q Max
#> -13.501 -3.409 0.005 3.311 12.616
#>
#> Coefficients:
#> Estimate Std. Error t value Pr(>|t|)
#> (Intercept) -12.3972 1.4459 -8.57 2.2e-16 ***
#> motivation -0.4372 0.0355 -12.31 < 2e-16 ***
#> ---
#> Signif. codes:
#> 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 4.87 on 398 degrees of freedom
#> Multiple R-squared: 0.276, Adjusted R-squared: 0.274
#> F-statistic: 151 on 1 and 398 DF, p-value: <2e-16
```

Il risultato prodotto dal modello di regressione è sbagliato: come conseguenza dell'errore di specificazione, il segno del coefficiente parziale di regressione della variabile “motivazione” è negativo, anche se nel vero modello di regressione tale coefficiente ha il segno opposto. Quindi, se noi interpretassimo il coefficiente parziale ottenuto in termini casuali, saremmo portati a concludere che la motivazione fa diminuire la prestazione anche se, in realtà (nel modello generatore dei dati), è vero l'opposto.

È facile vedere perché questo si verifica. Supponiamo che il vero modello sia

$$y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

che verrebbe stimato con

$$y = a + b_1 X_1 + b_2 X_2 + e.$$

Supponiamo che il ricercatore creda invece che

$$y = \alpha' + \beta'_1 X_1 + \varepsilon'$$

e quindi stimi

$$y = a' + b'_1 X_1 + e'$$

omettendo X_2 dal modello. Ci chiediamo che relazione ci sia tra b'_1 e b_1 .

La formula per b'_1 è

$$\begin{aligned} b'_1 &= \frac{\text{Cov}(X_1, Y)}{\text{Var}(X_1)} \\ &= \frac{\text{Cov}(X_1, a + b_1 X_1 + b_2 X_2 + e)}{\text{Var}(X_1)} \\ &= \frac{\text{Cov}(X_1, a) + b_1 \text{Cov}(X_1, X_1) + b_2 \text{Cov}(X_1, X_2) + \text{Cov}(X_1, e)}{\text{Var}(X_1)} \\ &= \frac{0 + b_1 \text{Var}(X_1) + b_2 \text{Cov}(X_1, X_2) + 0}{\text{Var}(X_1)} \\ &= b_1 + b_2 \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_1)}. \end{aligned}$$

Quindi, se X_2 viene erroneamente omissso dal modello, abbiamo che

$$\mathbb{E}(b'_1) = \beta_1 + \beta_2 \frac{\sigma_{12}}{\sigma_1^2}.$$

Verifichiamo per i dati dell'esempio che stiamo discutendo. Nel caso presente, X_1 è `motivation` e X_2 è `anxiety`. Dunque, applicando la formula precedente, otteniamo lo stesso valore per il coefficiente di regressione associato a `motivation` che era stato ottenuto adattando ai dati il modello `performance ~ motivation`, ovvero:

```
fm1$coef[2] + fm1$coef[3] *
  cov(sim_dat2$motivation, sim_dat2$anxiety) /
  var(sim_dat2$motivation)
#> motivation
#>      -0.4372
```

In conclusione, b'_1 è uno stimatore distorto di β_1 . Si noti che questa distorsione non scomparirà all'aumentare della numerosità campionaria, il che (in termini statistici) ci porta a concludere che un tale stimatore è *inconsistente*. Quello che succede in pratica è che alla variabile X_1 vengono attribuiti gli effetti delle variabili che sono state omesse dal modello. Si noti che una tale distorsione sistematica di b'_1 può essere evitata solo se si verificano due condizioni:

- $\beta_2 = 0$. Questo è ovvio, dato che, se $\beta_2 = 0$, ciò significa che il modello non è specificato in modo errato, cioè X_2 non appartiene al modello perché non ha un effetto diretto sulla Y .
- $\sigma_{12} = 0$. Cioè, se X_1 e X_2 sono incorrelate, allora l'omissione di una delle due variabili non comporta stime distorte dell'effetto dell'altra.

1.2.4 Soppressione

Le conseguenze dell'errore di specificazione sono chiamate “soppressione” (*suppression*). In generale, si ha soppressione quando (1) il valore assoluto del peso beta di un predittore è maggiore di quello della sua correlazione bivariata con il criterio o (2) i due hanno segni opposti.

- L'esempio descritto sopra è un caso di *soppressione negativa*, dove il predittore ha correlazioni bivariate positive con il criterio, ma si riceve un peso beta negativo nell'analisi di regressione multipla.

- Un secondo tipo di soppressione è la *soppressione classica*, in cui un predittore non è correlato al criterio ma riceve un peso beta diverso da zero se un altro predittore viene controllato.
- C'è anche la *soppressione reciproca* che può verificarsi quando due variabili sono correlate positivamente con il criterio ma negativamente tra loro.

1.2.5 Stepwise regression

Un'implicazione della soppressione è che i predittori non dovrebbero essere selezionati in base ai valori delle correlazioni bivariate con il criterio. Queste associazioni di ordine zero non controllano gli effetti degli altri predittori, quindi i loro valori possono essere fuorvianti rispetto ai coefficienti di regressione parziale per le stesse variabili. Per lo stesso motivo, il fatto che le correlazioni bivariate con il criterio siano statisticamente significative o meno è irrilevante per quanto riguarda la selezione dei predittori. Sebbene le procedure informatiche di regressione rendano facile tali processi di selezione dei predittori, i ricercatori dovrebbero evitare di usare tali metodi. Il rischio è che anche piccole, ma non rilevate, non-linearità o effetti indiretti tra i predittori possano seriamente distorcere i coefficienti di regressione parziale. È meglio selezionare giudiziosamente il minor numero di predittori sulla base di ragioni teoriche o dei risultati di ricerche precedenti.

Una volta selezionati, ci sono due modi di base per inserire i predittori nell'equazione di regressione: uno consiste nell'inserire tutti i predittori contemporaneamente. L'altro è inserirli nel corso di una serie di passaggi, ovvero mediante usando una procedura sequenziale. L'ordine di ingresso può essere determinato in base a uno di due diversi standard: teorici (razionali) o empirici (statistici). Lo standard razionale corrisponde alla regressione gerarchica, in cui si comunica al computer un ordine fisso per inserire i predittori. Ad esempio, a volte le variabili demografiche vengono inserite nel primo passaggio, quindi nel secondo passaggio viene inserita una variabile psicologica di interesse. Questo ordine non solo controlla le variabili demografiche ma permette anche di valutare il potere predittivo della variabile psicologica, al di là di quello delle semplici variabili demografiche. Quest'ultimo può essere stimato come l'aumento della correlazione multipla al quadrato, o ΔR^2 , da quella della fase 1 con solo predittori demografici a quella della fase 2 con tutti i predittori nell'equazione di regressione.

Un esempio di standard statistico è la regressione *stepwise*, in cui il computer seleziona l'inserimento dei predittori in base esclusivamente alla significatività statistica; cioè, viene chiesto: quale predittore, se inserito nell'equazione, avrebbe il valore p più piccolo per il test del suo coefficiente di regressione parziale? Dopo la selezione, i predittori in una fase successiva possono essere rimossi dall'equazione di regressione in base ai loro valori- p (ad esempio, se $p \geq .05$). Il processo *stepwise* si interrompe quando, aggiungendo più predittori, ΔR^2 non migliora. Varianti della regressione *stepwise* includono *forward inclusion*, in cui i predittori selezionati non vengono successivamente rimossi dal modello, e *backward elimination*, che inizia con tutti i predittori nel modello per poi rimuoverne alcuni in passi successivi. I problemi relativi ai metodi *stepwise* sono così gravi da essere effettivamente banditi in alcuni giornali. Un problema è che fanno leva su risultati che si ottengono per caso, in dipendenza delle idiosincrasie del campione (quindi, non replicabili).

In secondo luogo, una volta che un insieme finale di predittori selezionati razionalmente è stato inserito nell'equazione di regressione, tali predittori non dovrebbero essere successivamente rimossi se i loro coefficienti di regressione non sono statisticamente significativi: il ricercatore non dovrebbe sentirsi in dovere di lasciar perdere ogni predittore che non risulta statisticamente significativo. In campioni piccoli, la potenza dei test di significatività è bassa e la rimozione di un predittore non significativo può alterare sostanzialmente la soluzione. Se c'è una buona ragione per includere un predittore, allora è meglio lasciarlo nel modello, fino a prova contraria.

2

Invarianza di misura

I precedenti esempi di CFA presentati in questa dispensa sono stati stimati all'interno di un singolo gruppo, hanno utilizzato come input un'unica matrice covarianza e hanno portato alla stima dei parametri del modello sui quali non è stata imposta alcuna restrizione. In questo capitolo, le analisi precedenti verranno estese considerando il problema dell'invarianza di misura. Quello che ci chiediamo è se sia sensato considerare la medesima struttura fattoriale in gruppi diversi. In altre parole, ci chiediamo se viene misurata la stessa variabile latente tra gruppi diversi. Questa proprietà è chiamata *invarianza di misura* (Meredith, 1993). L'approccio che viene utilizzato per stabilire l'evidenze dell'invarianza di misura è attraverso l'analisi fattoriale confermativa a gruppi multipli (*multiple-group confirmatory factor analysis*, MG-CFA). È solo nella misura in cui viene dimostrata l'equivalenza di misura che sono possibili i confronti tra gruppi. Nel presente capitolo, verrà affrontato il problema di invarianza di misura considerando prima il caso di indicatori continui e poi il caso di indicatori categoriali.

2.1 Indicatori continui

2.1.1 Intercette degli item

In generale, i modelli di equazioni strutturali vengono utilizzati per modellare unicamente la matrice di covarianza delle variabili osservate in un set di dati. Ricordiamo che, quando abbiamo introdotto il modello dell'analisi fattoriale,

$$y_i = \mu + \lambda_j \xi_k + \delta_i,$$

per semplicità abbiamo ignorato la media μ degli indicatori esprimendo i dati osservati nei termini degli scarti dalla media, $y_i - \mu$, in quanto

ciò lascia immutate le covarianze. Tuttavia, in alcune applicazioni (quali l'invarianza di misura, ad esempio), è utile considerare anche le medie delle variabili osservate. Per ottenere ciò è possibile fare esplicito riferimento alle intercette delle equazioni precedenti nella sintassi `lavaan` del modello. È possibile fare riferimento all'intercetta di una variabile manifesta nel modo seguente:

```
variable ~ 1
```

laddove la parte sinistra dell'espressione precedente contiene il nome della variabile osservata a cui si fa riferimento. Per esempio, nel caso di un modello a due fattori comuni, è possibile aggiungere le intercette delle variabili manifeste nel modo seguente:

```
mod <- "  
  # two-factor model  
  f1 =~ x1 + x2 + x3  
  f2 =~ x4 + x5 + x6  
  # intercepts  
  x1 ~ 1  
  x2 ~ 1  
  x3 ~ 1  
  x4 ~ 1  
  x5 ~ 1  
  x6 ~ 1  
"
```

Tuttavia, è più conveniente omettere le intercette nella specificazione del modello e aggiungere l'argomento `meanstructure = TRUE` nella funzione `cfa()`:

```
fit <- cfa(  
  mod,  
  data = d,  
  meanstructure = TRUE  
)
```

Si noti che la statistica chi-quadrato e il numero di gradi di libertà sono gli stessi in un modello con o senza `meanstructure`. Il motivo è che, nel

caso di un modello con `meanstructure`, vengono introdotti nuovi dati (il valore della media per ciascuna delle p variabili osservate) ma vengono anche aggiunti p parametri da stimare (un'intercetta per ciascuna delle p variabili osservate). Il risultato finale è che la bontà dell'adattamento resta immutata. In pratica, l'unico motivo per aggiungere le intercette nella sintassi del modello è quello di stabilire dei vincoli nella stima di tali parametri.

2.1.2 Terminologia

La discussione dell'invarianza di misura nel contesto della CFA fa uso della seguente terminologia.

- L'invarianza configurale (*configural invariance*) verifica se la struttura dei fattori è la stessa tra i gruppi, ovvero verifica la presenza dello stesso numero di fattori e di pattern di saturazioni fattoriali simili tra gruppi.
- L'invarianza metrica (*metric invariance*) o “invarianza fattoriale debole” (*weak factorial invariance*) verifica se i carichi fattoriali degli elementi sono gli stessi tra i gruppi. Invarianza scalare (*scalar invariance*) o “invarianza fattoriale forte” verifica se le intercette sono le stesse tra i gruppi.
- L'invarianza fattoriale rigorosa (*strict factorial invariance*), infine, verifica se i residui degli indicatori sono gli stessi tra i gruppi.

2.1.3 Un esempio concreto

Consideriamo qui un esempio discusso da [Brown \(2015\)](#). Il modello CFA riguarda un modello di misurazione per la depressione maggiore così come definita nel DSM-IV. Ci sono 9 indicatori:

- MDD1, depressed mood;
- MDD2, loss of interest in usual activities;
- MDD3, weight/appetite change;
- MDD4, sleep disturbance;
- MDD5, psychomotor agitation/retardation;
- MDD6, fatigue/loss of energy;
- MDD7, feelings of worthlessness/guilt;
- MDD8, concentration difficulties;
- MDD9, thoughts of death/suicidality.

Leggiamo i dati:

```
d <- readRDS(  
  here::here("data", "mdd_sex.RDS")  
)
```

Il problema riguarda l'invarianza fattoriale in funzione del genere. Consideriamo il seguente modello:

```
model_mdd <- "  
  MDD =~ mdd1 + mdd2 + mdd3 + mdd4 + mdd5 + mdd6 + mdd7 + mdd8 +  
          mdd9  
  mdd1 ~~ mdd2  
"
```

Si noti la presenza di una correlazione residua tra gli indicatori `mdd1` e `mdd2`.

Esaminiamo dunque le varie forme di invarianza fattoriale:

```
# configural invariance  
fit_ef <- cfa(  
  model_mdd,  
  data = d,  
  group = "sex",  
  meanstructure = TRUE  
)  
# equal factor loadings  
fit_efl <- update(  
  fit_ef,  
  group.equal = c("loadings")  
)  
# equal indicator intercepts  
fit_eii <- update(  
  fit_efl,  
  group.equal = c("loadings", "intercepts")  
)  
# equal indicator error variances  
fit_eir <- update(  
  fit_eii,
```

```

    group.equal = c("loadings", "intercepts", "residuals")
  )
  # equal factor variances
  fit_fv <- update(
    fit_eir,
    group.equal = c(
      "loadings", "intercepts", "residuals",
      "lv.variances"
    )
  )
  # equal latent means
  fit_fm <- update(
    fit_fv,
    group.equal = c(
      "loadings", "intercepts", "residuals",
      "lv.variances", "means"
    )
  )
)

```

Confrontiamo i modelli:

```

lavTestLRT(fit_ef, fit_efl, fit_eii, fit_eir, fit_fv, fit_fm)
#> Chi-Squared Difference Test
#>
#>      Df   AIC   BIC Chisq Chisq diff Df diff
#> fit_ef  52 27526 27785  98.9
#> fit_efl 60 27514 27735 102.8      3.93      8
#> fit_eii 68 27510 27695 115.3     12.47      8
#> fit_eir 77 27502 27645 125.0      9.71      9
#> fit_fv  78 27501 27639 125.8      0.79      1
#> fit_fm  79 27501 27635 127.7      1.92      1
#>      Pr(>Chisq)
#> fit_ef
#> fit_efl      0.86
#> fit_eii      0.13
#> fit_eir      0.37
#> fit_fv      0.37
#> fit_fm      0.17

```

Il confronto tra i precedenti modelli nidificati che introducono vincoli sempre più tringenti sui parametri indica che non vi è una “significativa” perdita di bontà dell’adattamento passando dal modello congenerico al modello che assume l’uguaglianza delle saturazioni fattoriali, delle inter-cette, delle varianze residue, delle varianze delle variabili latenti e delle medie nei due gruppi. Questi dati, dunque, forniscono forti evidenze di invarianza fattoriale tra maschi e femmine in relazione al costrutto di depressione maggiore.

2.2 Variabili a livello ordinale

I test di invarianza per i dati ordinali sono diversi da quelli utilizzati con continuo variabili in termini dello stimatore che viene utilizzato e del tipo di analisi statistica che viene svolta. Le variabili ordinali hanno opzioni di risposta che hanno un ordine logico, come come la gamma da fortemente in disaccordo a fortemente d’accordo; o da mai, a volte, spesso, a sempre. Queste opzioni possono essere ordinate logicamente e per convenzione ad esse venono assegnati valori numerici interi. Tuttavia, poiché le risposte ordinali non corrispondono a veri valori quantitativi (come 0 volte a settimana, 5 volte a settimana e 10 volte a settimana), l’assegnazione di numeri alle risposte ordinali da utilizzare nelle analisi è arbitraria. Ad esempio, le stesse cinque opzioni di risposta ordinate potrebbero essere assegnati valori da 0 a 4, da 1 a 5 o da 5 a 1. Pertanto, i dati ordinali non possono essere analizzati come se fossero continui. “Le medie, le varianze e le covarianze delle variabili ordinali non hanno significato” (Jöreskog, p. 1). Tuttavia, l’ordinamento delle risposte può essere analizzato tramite apposite procedure specifiche ai dati ordinali. Abbiamo visto in precedenza come lo stimatore attualmente consigliato per i dati ordinali è quello dei minimi quadrati ponderati. Consideriamo qui il problema dell’invarianza di misura.

Oltre alla stima WLS, nell’analisi di dati ordinali viene utilizzata una matrice di correlazione policorica invece della solita matrice di covarianza. Per ciascuna coppia di variabili ordinali viene calcolata una correlazione policorica sulla base dell’ipotesi che una variabile continua latente normalmente distribuita sia responsabile delle frequenze osservate delle risposte ordinali in ciascuna variabile. Secondo questa ipotesi, ogni valore di risposta ordinale osservato corrisponde a un intervallo di valori nor-

malizzati tra soglie, o cutoff, sulla variabile continua latente sottostante. Tali soglie ($\tau_1, \tau_2, \dots, \tau_k$) sono dei margini verticali che suddividono l'area sottesa alla funzione di densità della distribuzione normale sottostante in k sezioni, ciascuna delle quali corrisponde alla frequenza del punteggio ordinale che è stato osservato in quella categoria di risposta. L'invarianza delle soglie (*threshold invariance*) assume che tali soglie necessarie per definire le correlazioni policoriche siano invarianti tra gruppi.

Wu and Estabrook (2016) mostrano come la procedura per la valutazione dell'invarianza di misura che è stata descritta in precedenza debba essere modificata nel caso di indicatori categoriali. La procedura usuale consiste nel definire prima un modello di riferimento e successivamente di imporre restrizioni crescenti ai parametri. Secondo Wu and Estabrook (2016), tale approccio non è ottimale nel caso di dati categoriali perché dipende fortemente dal modo in cui vengono definite le soglie rispetto alle scale delle risposte continue latenti nel modello di base. Dunque, secondo Wu and Estabrook (2016), è prima necessario valutare l'equivalenza delle soglie tra gruppi (*threshold model*) e poi valutare il modello che ipotizza l'equivalenza tra i gruppi delle saturazioni fattoriali.

Per illustrare tale procedura, replichiamo qui il tutorial messo a punto da Svetina et al. (2020). Questi autori utilizzano quattro item di una scala del bullismo per tre paesi (31 = Azerbaigian; 40 = Austria; 246 = Finlandia). Tutti gli item sono misurati su una scala di tipo Likert a 4 punti, che va da 0 (mai) a 3 (almeno una volta alla settimana). Gli item di questa scala chiedono agli studenti quante volte durante l'anno è successo loro a scuola uno degli episodi in questione; per esempio: “mi prendevano in giro o mi insultavano”. Le dimensioni del campione per l'Azerbaigian, l'Austria e la Finlandia sono rispettivamente 3808, 4457 e 4520. Leggiamo in dati in R:

```
dat <- read.table("data/BULLY.dat", header = FALSE)
names(dat) <- c("IDCNTRY", "R09A", "R09B", "R09C", "R09D")
head(dat)
#>   IDCNTRY R09A R09B R09C R09D
#> 1      31    3    3    0    0
#> 2      31    0    0    0    0
#> 3      31    3    2    1    3
#> 4      31    0    0    3    0
```

```
#> 5      31      0      0      0      0
#> 6      31      0      0      0      0
```

Viene creata la matrice `all.results` per immagazzinare i risultati dei diversi modelli che verranno confrontati, chiamati *baseline* (nessun vincolo tra gruppi), *proposition 4* (equivalenza delle soglie tra gruppi), e *proposition 7* (equivalenza delle soglie e delle saturazioni fattoriali tra gruppi). Gli indici di bontà dell'adattamento che verranno considerati sono: chi-square, df, p, RMSEA, CFI, e TLI.

```
all.results <- matrix(NA, ncol = 6, nrow = 3)
```

2.2.1 Baseline model

Nel *baseline model* non viene posto alcun vincolo tra i gruppi o le misure ripetute:

```
mod.cat <- "F1 =~ R09A + R09B + R09C + R09D"
```

```
baseline <- measEq.syntax(
  configural.model = mod.cat,
  data = dat,
  ordered = c("R09A", "R09B", "R09C", "R09D"),
  parameterization = "delta",
  ID.fac = "std.lv",
  ID.cat = "Wu.Estabrook.2016",
  group = "IDCNTRY",
  group.equal = "configural"
)
```

Informazioni sul modello baseline si ottengono nel modo seguente:

```
summary(baseline)
```

Le proprietà del modello possono essere esplicitate con la seguente istruzione:

```
cat(as.character(baseline))
```

Per potere essere passato a `lavaan`, l'oggetto `baseline` deve essere in formato `char`:

```
model.baseline <- as.character(baseline)
```

Adattiamo il modello ai dati:

```
fit.baseline <- cfa(
  model.baseline,
  data = dat,
  group = "IDCNTRY",
  ordered = c("R09A", "R09B", "R09C", "R09D")
)
```

Salviamo i risultati:

```
all.results[1, ] <-
  round(data.matrix(
    fitmeasures(fit.baseline, fit.measures = c(
      "chisq.scaled", "df.scaled", "pvalue.scaled",
      "rmsea.scaled", "cfi.scaled", "tli.scaled"
    ))
  ),
  digits = 3
)
```

2.2.2 Invarianza delle soglie

Consideriamo ora il modello *threshold invariance* (chiamata *Proposition 4* da [Wu and Estabrook, 2016](#)).

```
prop4 <- measEq.syntax(
  configural.model = mod.cat,
  data = dat,
  ordered = c("R09A", "R09B", "R09C", "R09D"),
```

```

parameterization = "delta",
ID.fac = "std.lv",
ID.cat = "Wu.Estabrook.2016",
group = "IDCNTRY",
group.equal = c("thresholds")
)

```

Adattiamo il modello ai dati:

```

model.prop4 <- as.character(prop4)
fit.prop4 <- cfa(
  model.prop4,
  data = dat,
  group = "IDCNTRY",
  ordered = c("R09A", "R09B", "R09C", "R09D")
)

```

Salviamo i risultati

```

# store model fit information for proposition 4
all.results[2, ] <-
  round(data.matrix(
    fitmeasures(fit.prop4, fit.measures = c(
      "chisq.scaled", "df.scaled", "pvalue.scaled",
      "rmsea.scaled", "cfi.scaled", "tli.scaled"
    ))
  ),
  digits = 3
)

```

Eseguiamo il confronto tra il modello di *threshold invariance* e il modello di *baseline*:

```

lavTestLRT(fit.baseline, fit.prop4)
#> Scaled Chi-Squared Difference Test (method = "satorra.2000")
#>
#> lavaan NOTE:
#>   The "Chisq" column contains standard test statistics, not the

```

```
#>      robust test that should be reported per model. A robust difference
#>      test is a function of two standard (not robust) statistics.
#>
#>              Df AIC BIC Chisq Chisq diff Df diff
#> fit.baseline  6      26.9
#> fit.prop4    14      42.2      46.1      8
#>              Pr(>Chisq)
#> fit.baseline
#> fit.prop4      2.2e-07 ***
#> ---
#> Signif. codes:
#> 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.2.3 Threshold and loading invariance

Consideriamo ora il modello *threshold and loading invariance* (chiamato *Proposition 7* da [Wu and Estabrook, 2016](#)).

```
prop7 <- measEq.syntax(
  configural.model = mod.cat,
  data = dat,
  ordered = c("R09A", "R09B", "R09C", "R09D"),
  parameterization = "delta",
  ID.fac = "std.lv",
  ID.cat = "Wu.Estabrook.2016",
  group = "IDCNTRY",
  group.equal = c("thresholds", "loadings")
)
```

Adattiamo il modello ai dati:

```
model.prop7 <- as.character(prop7)
fit.prop7 <- cfa(
  model.prop7,
  data = dat, group = "IDCNTRY",
  ordered = c("R09A", "R09B", "R09C", "R09D")
)
```

Salviamo i risultati:

```
all.results[3, ] <-
  round(data.matrix(
    fitmeasures(fit.prop7, fit.measures = c(
      "chisq.scaled", "df.scaled", "pvalue.scaled",
      "rmsea.scaled", "cfi.scaled", "tli.scaled"
    ))
  ), digits = 3)

column.names <-
  c(
    "chisq.scaled", "df.scaled", "pvalue.scaled", "rmsea.scaled",
    "cfi.scaled", "tli.scaled"
  )

row.names <- c("baseline", "prop4", "prop7")

colnames(all.results) <- column.names
rownames(all.results) <- row.names
```

Eseguiamo i confronti tra modelli:

```
lavTestLRT(fit.prop4, fit.prop7)
#> Scaled Chi-Squared Difference Test (method = "satorra.2000")
#>
#> lavaan NOTE:
#>   The "Chisq" column contains standard test statistics, not the
#>   robust test that should be reported per model. A robust difference
#>   test is a function of two standard (not robust) statistics.
#>
#>           Df AIC BIC Chisq Chisq diff Df diff
#> fit.prop4 14      42.2
#> fit.prop7 20      93.1      54.7      6
#>           Pr(>Chisq)
#> fit.prop4
#> fit.prop7 5.4e-10 ***
#> ---
```

```
#> Signif. codes:
#> 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

lavTestLRT(fit.prop7, fit.baseline)
#> Scaled Chi-Squared Difference Test (method = "satorra.2000")
#>
#> lavaan NOTE:
#>   The "Chisq" column contains standard test statistics, not the
#>   robust test that should be reported per model. A robust difference
#>   test is a function of two standard (not robust) statistics.
#>
#>           Df AIC BIC Chisq Chisq diff Df diff
#> fit.baseline 6      26.9
#> fit.prop7    20      93.1      102      14
#>           Pr(>Chisq)
#> fit.baseline
#> fit.prop7      2.3e-15 ***
#> ---
#> Signif. codes:
#> 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Un confronto tra gli indici di bontà di adattamento dei tre modelli è fornito di seguito:

```
all.results
#>           chisq.scaled df.scaled pvalue.scaled
#> baseline           50.94           6           0
#> prop4             111.98          14           0
#> prop7             210.64          20           0
#>           rmsea.scaled cfi.scaled tli.scaled
#> baseline           0.042           0.997           0.991
#> prop4              0.041           0.993           0.992
#> prop7              0.047           0.987           0.989
```

In conclusione, nel caso presente, il test del rapporto di verosimiglianza indica che non viene rispettata neppure l'invarianza delle soglie tra gruppi. Gli altri confronti, dunque, sono superflui.



Bibliografia

- Beaujean, A. A. (2014). *Latent variable modeling using R: A step-by-step guide*. Routledge.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.
- Lord, F. M. and Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.
- Mauro, R. (1990). Understanding LOVE (left out variables error): A method for estimating the effects of omitted variables. *Psychological Bulletin*, 108(2):314–329.
- McDonald, R. P. (2013). *Test theory: A unified treatment*. Psychology Press.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4):525–543.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Svetina, D., Rutkowski, L., and Rutkowski, D. (2020). Multiple-group invariance with categorical outcomes using updated guidelines: an illustration using m plus and the lavaan/semtools packages. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(1):111–130.
- Wu, H. and Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, 81(4):1014–1045.