

Mixture SGR:
**un nuovo paradigma nella gestione delle
manipolazioni di dati self report**



Corso di Dottorato in: Scienze Psicologiche e della Formazione

Ciclo XXX

Settore disciplinare: M-PSI/03 Psicometria

Dottorando: Marco Bressan

Tutor: Prof. Luigi Lombardi

Anno accademico 2017/18

Rovereto 2 Marzo 2018

Indice

| | | |
|----------|--|-----------|
| 1 | Introduzione | 1 |
| 1.1 | Il faking | 1 |
| 1.2 | SGR, Sample Generation by Replacement | 3 |
| 2 | L'effetto del faking sulla correlazione di due item ordinali | 11 |
| 2.1 | Introduzione | 11 |
| 2.2 | Due misure di correlazione | 14 |
| 2.2.1 | Indice di correlazione di Pearson | 15 |
| 2.2.2 | Indice di correlazione di Spearman | 16 |
| 2.3 | SGR tradizionale | 17 |
| 2.4 | Mixture SGR | 19 |
| 2.4.1 | Modello di risposta onesta | 19 |
| 2.4.2 | Modelli di manipolazione fake | 20 |
| 2.5 | Valutazione a livello di popolazione | 21 |
| 2.5.1 | L'indice Q di Cohen | 21 |
| 2.6 | Risultati dell'analisi a livello di popolazione | 23 |
| 2.7 | Studio simulativo | 26 |
| 2.7.1 | Disegno della simulazione | 27 |
| 2.8 | Risultati | 28 |
| 2.9 | Breve discussione e conclusioni | 32 |
| 3 | Identificare pattern di risposte manipolate | 39 |
| 3.1 | Introduzione | 39 |
| 3.2 | Caratteristiche delle perturbazioni dei pattern in simulazioni Monte Carlo | 40 |
| 3.2.1 | Codifica dicotomica della risposta | 41 |
| 3.2.2 | Codifica politomica ordinale della risposta | 49 |
| 3.3 | Studio simulativo | 50 |
| 3.3.1 | Quattro indici di fit del soggetto | 51 |
| 3.3.2 | Modello di risposta onesta | 52 |

| | | |
|----------|--|-----------|
| 3.3.3 | Modelli di manipolazione fake | 53 |
| 3.3.4 | Il disegno sperimentale | 53 |
| 3.3.5 | Analisi dei dati | 56 |
| 3.4 | Analisi Likelihood pattern e dataset soggetti a faking | 61 |
| 3.4.1 | Analisi Likelihood locale | 63 |
| 3.4.2 | Analisi Likelihood globale | 64 |
| 3.5 | Conclusioni | 66 |
| 4 | Analisi formale dei modelli di mixture SGR | 69 |
| 4.1 | L'SGR a livello di popolazione, mixture SGR | 69 |
| 4.1.1 | Caso univariato | 70 |
| 4.1.2 | Caso bivariato | 72 |
| 4.2 | Derivazioni Analitiche | 74 |
| 4.2.1 | 1 item | 74 |
| 4.3 | 2 item | 75 |
| 5 | Discussione Generale | 77 |
| A | Calcolo parallelo | 81 |
| B | Codici di programmazione R | 85 |
| B.1 | Capitolo 2 | 85 |
| B.2 | Capitolo 3 | 89 |
| | Bibliografia | 97 |

Elenco delle figure

- 1.1 Rappresentazione dello schema SGR: l'individuo sottoposto ad uno stimolo, in una prima fase identifica la risposta onesta, d , e in una seconda elabora, e poi esprime, una risposta manipolata, f 4
- 1.2 Nella Figura sono rappresentati quattro esempio di distribuzione beta G con una variabile continua definita in $[a = 1, b = 7]$. Inoltre sono rappresentati i due tipi di discretizzazione senza, con i pallini vuoti, e con correzione, pallini pieni. 7
- 1.3 Nella Figura sono rappresentati quattro esempio di distribuzione di sostituzione con una variabile random discreta a 7 valori, valore non corrotto $h = 3$ e probabilità di sostituzione $\pi = .4$. Inoltre sono rappresentati i due tipi di discretizzazione senza, con i pallini vuoti, e con correzione, pallini pieni. 9
- 2.1 Quattro esempio di distribuzione di sostituzione per una variabile discreta a 7 livelli. Ogni colonna del grafico corrisponde a una differente distribuzione di sostituzione condizionata. Ogni riga del grafico corrisponde a uno dei 7 livelli discreti non manipolati. I parametri assegnati ai quattro modelli sono: modello *slight* ($\gamma = 1.5, \delta = 4$) e modello *extreme*: ($\gamma = 4, \delta = 1.5$). Le distribuzioni di sostituzione sono utilizzate nelle due configurazioni, rispettivamente, *faking good* e *faking bad* . . . 22
- 2.2 Differenza di correlazione $\Delta\rho = \rho_m - \rho_d$ come funzione della statistica Q di Cohen della correlazione originaria $\rho_d \in \{-0.9, -0.6, -0.3, 0, 0.3, 0.6, 0.9\}$. Le linee verticali identificano dimensioni degli effetti piccole, medie o grandi. 23

- 2.3 La statistica Q di Cohen come funzione della proporzione di fakers nella popolazione, modello di faking (slight, average, extreme) e della correlazione non corrotta. L'analisi a livello di popolazione riguarda i due indici di correlazione (Pearson e Spearman) con lo stesso modello e direzione di faking applicato su due variabili discrete con un uguale e limitato numero di possibili livelli ($v = 5$) e distribuzioni marginali con le seguenti masse di probabilità 0.06, 0.25, 0.38, 0.25, e 0.06. (Il modello average, $\gamma = 3, \delta = 3$, rappresenta un compromesso tra il modello slight e quello extreme) 24
- 2.4 La statistica Q di Cohen come funzione della proporzione di fakers nella popolazione, modello di faking (slight, average, extreme) e della correlazione non corrotta. L'analisi a livello di popolazione riguarda i due indici di correlazione (Pearson e Spearman) sotto lo stesso modello e opposta direzione di faking applicato su due variabili discrete con un uguale e limitato numero di possibili livelli ($v = 5$) e distribuzioni marginali con le seguenti masse di probabilità 0.06, 0.25, 0.38, 0.25, e 0.06. (Il modello average, $\gamma = 3, \delta = 3$, rappresenta un compromesso tra il modello slight e quello extreme) 25
- 2.5 La statistica Q di Cohen come funzione della proporzione di fakers e modello di faking. In questo esempio, sono mostrati $\rho_d = 0$, $v = 5$ e stessa direzione di faking. 33
- 2.6 La statistica Q di Cohen come funzione della proporzione di fakers e direzione di faking nelle due variabili. In questo esempio, sono mostrati $\rho_d = 0$, $v = 5$ e un modello di faking extreme. 33
- 2.7 La statistica Q di Cohen come funzione della proporzione di fakers e 3 numerosità dei livelli delle variabili discrete ($v = 2, 5, 7$). In questo esempio, sono mostrati $\rho_d = 0$, un modello di faking extreme con la stessa direzione. Nota: nel primo riquadro ($v = 2$) i due indici di correlazione sono completamente sovrapposti. 34
- 2.8 La statistica Q di Cohen come funzione della proporzione di fakers e la correlazione non corrotta a 5 livelli ($\rho_d = -0.9, -0.3, 0, 0.3, 0.9$). In questo esempio, sono mostrati $v = 5$, modello di faking extreme con la stessa direzione. 34

| | | |
|------|---|----|
| 2.9 | La statistica Q di Cohen come funzione della proporzione di fakers e dimensione campionaria in 4 livelli ($n = 25, 50, 100, 1000$). In questo esempio, sono mostrati $\rho_d = 0$, $v = 5$, e un modello di faking extreme con la stessa direzione. A.P. e A.S., rispettivamente, stanno per correlazione di Pearson oppure Spearman calcolate a livello di popolazione. S.P. e S.S., stanno, rispettivamente, per correlazione di Pearson e correlazione di Spearman stimate tramite procedura Monte Carlo (MC). | 35 |
| 2.10 | Boxplot della statistica campionaria q come funzione della proporzione di fakers nel campione e della dimensione del campione a 4 livelli ($n = 20, 50, 100, 1000$). In questo esempio, è stata calcolato solo q di Spearman; sono mostrati $\rho_d = 0$, $v = 5$, e un modello di faking extreme con la stessa direzione. | 36 |
| 3.1 | Percentuale di veri positivi sul totale di soggetti segnalati come funzione della probabilità di aberrazione di una risposta, il modello di faking e la percentuale di fakers nel campione, tenendo costanti il numero di livelli dell'item (5), il numero di item (5) e la dimensione campionaria (300). . | 57 |
| 3.2 | Percentuale di veri positivi sul totale di soggetti segnalati come funzione della probabilità di aberrazione di una risposta, il modello di faking e la percentuale di fakers nel campione, tenendo costanti il numero di livelli dell'item (7), il numero di item (5) e la dimensione campionaria (300). . | 58 |
| 3.3 | Percentuale di veri positivi sul totale di soggetti segnalati come funzione della probabilità di aberrazione di una risposta, il modello di faking e la percentuale di fakers nel campione, tenendo costanti il numero di livelli dell'item (5), il numero di item (20) e la dimensione campionaria (300). . | 59 |
| 3.4 | Percentuale di veri positivi sul totale di soggetti segnalati come funzione della probabilità di aberrazione di una risposta, il modello di faking e la percentuale di fakers nel campione, tenendo costanti il numero di livelli dell'item (7), il numero di item (20) e la dimensione campionaria (300). . | 60 |
| 3.5 | Distribuzioni delle $\log \mathcal{L}_l$ di pattern generati da modello, in nero, completamente aberrati, in rosso, e casuali, in verde | 64 |
| 3.6 | Distribuzioni delle $\log \mathcal{L}_g$ di matrici generate da modello, in nero, completamente aberrata, in rosso, e con percentuali di fakers, in grigio . . . | 65 |
| 3.7 | Distribuzioni delle $\log \mathcal{L}_g$ di matrici generate da modello, in nero, completamente aberrate, in rosso, e casuali, in verde | 66 |

Elenco delle tabelle

| | | |
|-----|---|----|
| 2.1 | Dimensione dell'effetto q come funzione della percentuale di fakers, correlazione originale e tipo di modello di faking. Caso di variabili binarie, $V = 2$, con la stessa direzione di faking. | 29 |
| 2.2 | Dimensione dell'effetto q come funzione della percentuale di fakers, correlazione originale e tipo di modello di faking. Caso di variabili binarie, $V = 5$, con la stessa direzione di faking. | 30 |
| 2.3 | Dimensione dell'effetto q come funzione della percentuale di fakers, correlazione originale e tipo di modello di faking. Caso di variabili binarie, $V = 5$, con la stessa direzione di faking. | 31 |
| 3.1 | Manipolazioni per dati dicotomici descritte in letteratura utilizzate in simulazioni Monte Carlo. θ rappresenta l'abilità del soggetto in un modello IRT; v è il numero di possibili risposte all'item. | 45 |
| 3.2 | La classificazione delle diverse aberrazioni definita dalle 5 caratteristiche identificate. | 47 |
| 3.3 | Procedure per imitare alcune manipolazioni dicotomiche attraverso la procedura SGR. Il termine o rappresenta la proporzione di risposte da sostituire e c rappresenta la probabilità di risposta corretta nella nuova estrazione. π^+ è la proporzione di celle da sostituire con una risposta maggiore all'onesta, π^- è la proporzione di celle da sostituire con una risposta minore all'onesta. | 49 |
| 3.4 | Soglie dei 12 item del questionario AEP/A | 63 |
| 3.5 | Correlazione tra i 12 item del questionario AEP/A | 63 |

Glossario della simbologia

| Simbolo | Funzione |
|---|---------------------------------------|
| a, b, c, \dots, y, z | Numero scalare. |
| $\mathbf{a}, \mathbf{b}, \mathbf{c}, \dots, \mathbf{y}, \mathbf{z}$ | Vettore con una dimensione. |
| A, B, C, \dots, Y, Z | Variabile casuale. |
| $\mathbf{A}, \mathbf{B}, \mathbf{C}, \dots, \mathbf{Y}, \mathbf{Z}$ | Vettore con due dimensioni (matrice). |
| $\alpha, \beta, \gamma, \dots, \psi, \omega$ | Parametri scalari. |
| $\boldsymbol{\theta}$ | Vettore di parametri |
| $[a, b]$ | Intervallo chiuso. |
| $[\dots]$ | Definisce un vettore. |
| $\{$ | Sistema di equazioni. |
| $\vec{}$ | Funzione di vettorizzazione. |
| \times | Prodotto vettoriale. |
| \otimes | Moltiplicazione di Kronecker. |

Capitolo 1

Introduzione

1.1 Il faking

Il problema che sta alla base di questa tesi è il faking, l'atto di mentire a qualcuno mostrandosi intenzionalmente diversi da quello che si è, auto-dichiarando il falso, con l'obiettivo di ottenere un qualsiasi genere di guadagno (es., Furnham, 1986; McFarland & Ryan, 2000; Zickar & Robie, 1999; MacCann, Ziegler, & Roberts, 2012). Uno dei problemi più importanti associati alla presenza di manipolazioni di tipo faking è che nella maggior parte delle situazioni empiriche non si hanno sufficienti informazioni per essere certi circa la reale sincerità dei partecipanti nel rispondere agli item di questionari di tipo self-report, specialmente quando tali item risultano espliciti o trasparenti e, dunque, il rispondente può facilmente inferire quale sia l'oggetto di indagine del questionario (es, Griffith & Converse, 2011; Levin & Zickar, 2002; Rosse, Stecher, Miller, & Levin, 1998). In ogni caso, spesso ci si affida completamente alle dichiarazioni dei partecipanti e non si ha modo di poter accertarne la veridicità. La tendenza al mentire è spesso accentuata quando si indagano argomenti sensibili, privati oppure illegali. Un esempio di comportamento di faking, riprodotto nella vita reale, potrebbe verificarsi quando un candidato per un posto di lavoro distorce le proprie risposte ad un questionario di personalità cercando di aumentare le probabilità di essere assunto (es., Paulhus, 1984; Zickar & Robie, 1999; Donovan, Dwight, & Schneider, 2014).

Il comportamento di faking è un problema trasversale che riguarda la ricerca sul comportamento umano in molti contesti quali la psicologia (Hopwood, Talbert, Morey, & Rogers, 2008), le scienze sociali (Van der Geest & Sarkodie, 1998) ed economiche (Crawford, 2003), la psichiatria (Beaber, Marston, Michelli, & Mills, 1985), la medicina forense (Gray, MacCulloch, Smith, Morris, & Snowden, 2003) e le frodi scientifiche (Marshall, 2000). Esso è un problema importante sia perché va ad alterare l'affidabilità di un test quando lo si usa per le decisioni in ambito inferenziale, sia perché i dati

così raccolti perdono di validità e quindi le relazioni tra variabili misurate perdono di significato.

Diversi studi hanno dimostrato che istruendo a dovere i partecipanti a un questionario, essi sono in grado di modificare le risposte fornendo risposte più estreme (es, Furnham, 1986; Hesketh, Griffin, & Grayson, 2004; McFarland & Ryan, 2000; Viswesvaran & Ones, 1999).

Gli effetti conosciuti del faking da un punto di vista psicométrico agiscono sulla struttura della covarianza delle scale distorte. Generalmente si è visto che i punteggi che hanno subito una manipolazione da faking, in una situazione in cui i soggetti venivano motivati, tendono ad avere una varianza più piccola e una minore affidabilità (Eysenck, Eysenck, & Shaw, 1974; Topping & O’Gorman, 1997; Ellingson, Smith, & Sackett, 2001; Hesketh et al., 2004). Allo stesso tempo però è stato osservato che quando è stato chiesto ai soggetti di mentire intenzionalmente, questo aumenta il grado di covariazione tra gli item manipolati (Ellingson, Sackett, & Hough, 1999; Galić, Jerneić, & Kovačić, 2012; Pauls & Crost, 2005; Zickar & Robie, 1999; Ziegler & Buehner, 2009).

Generalmente, viene fatta una distinzione tra faking good e faking bad. Questa distinzione è funzione dell’obiettivo della menzogna: se il soggetto vuole mostrarsi "migliore" di quel che realmente è si parlerà di faking good, se al contrario, il soggetto cerca di spingere su alcuni propri "difetti" o "aspetti meno gradevoli" si parlerà di faking bad. Un esempio di faking good potrebbe essere il mostrarsi gentili e disponibili per poter ottenere la fiducia di qualcun altro, un esempio di faking bad è dichiarare di avere qualche disturbo psico-fisico per poter evitare qualche mansione sul posto di lavoro. Nella psicomètria, un esempio di faking good potrebbe essere la dichiarazione mendace di un tossicodipendente, che rivolgendosi a un centro di supporto contro le dipendenze, dichiara di non utilizzare droghe con l’obiettivo di farsi prescrivere farmaci psicoattivi dal medico. Un esempio di faking bad si può osservare quando nei questionari compilati per ottenere rimborsi finanziari dalle assicurazioni infortunistiche, gli sfortunati incidentati esagerano volontariamente (e fraudolentemente) i propri sintomi (Hall & Hall, 2007).

Come si esprimono in un questionario i comportamenti di faking good e faking bad dipende, ovviamente, dal significato di ogni singolo item, ma in tutta questa tesi gli item verranno considerati omogenei come significato, cioè più un punteggio è a destra, più è socialmente desiderabile (faking good), mentre un punteggio basso, a sinistra, rappresenta qualcosa di sgradevole (faking bad). Questa semplificazione non impatta assolutamente sulla qualità generale delle conclusioni. In una situazione pratica, con dati reali, basterebbe solamente rovesciare i punteggi degli item con un significato

inverso; operazione che si fa abitualmente con i dati grezzi prima di procedere all'analisi.

I ricercatori hanno cercato di arginare l'impatto delle manipolazioni da faking attraverso diversi metodi. Per esempio, sono stati utilizzati metodi etnografici (es. tecniche nominative e campionamenti a palla di neve) per indagare comportamenti usualmente stigmatizzati dall'opinione pubblica, dove la dichiarazione viene spesso manipolata volontariamente (Tracy & Fox, 1981; Miller, 1981). Similmente, metodi psicometrici sono stato sviluppati per identificare e valutare le risposte dei soggetti che fingono un disagio (faking bad, malingering) o che si mettono sulla difensiva (faking good, autoinganno, desiderabilità sociale) utilizzando approcci di analisi fattoriale (es, Paulhus, 1991; Zickar & Robie, 1999; Ferrando, 2005; J.-P. Fox & Meijer, 2008; Holden & Book, 2009; Ziegler & Buehner, 2009; Leite & Cooper, 2010; Ferrando & Anguiano-Carrasco, 2011), factor mixture models (Leite & Cooper, 2010) e tecniche di diagnostica di casi (Pek & MacCallum, 2011). Un'altra ben nota metodologia è rappresentata dalle risposte randomizzate (RR, Warner, 1965; J. A. Fox & Tracy, 1986; Chaudhuri & Mukerjee, 1988). Questa tecnica è un approccio generale che fu sviluppato nella comunità statistica con l'obiettivo di proteggere la riservatezza ed è stato utilizzato specialmente in questionari auto-somministrati su larga scala (es, Campbell, 1987; J. E. Cohen, 1987; Kolata, 1987). L'approccio RR è caratterizzato da un complesso processo di campionamento e dalla necessità di una grande numero di soggetti per rendere le stime sufficientemente attendibili. L'approccio RR è stato talvolta criticato sia perché i questionari così costruiti non sono di facile compilazione da parte dei rispondenti, sia perché vengono chieste molte informazioni che possono sembrare inutili al compilatore, oppure sembrare inganni da parte del ricercatore (Campbell, 1987).

1.2 SGR, Sample Generation by Replacement

In questa sezione descriveremo brevemente un approccio probabilistico, chiamato Sample Generation by Replacement (SGR; Lombardi & Pastore, 2012; Pastore & Lombardi, 2014; Lombardi, Pastore, Nucci, & Bobbio, 2014), che costituirà il modello teorico di riferimento utilizzato in questa tesi per simulare fake data (ovvero dati le cui osservazioni sono state perturbate da processi di faking) e valutare gli effetti del faking sui risultati di vari indici statistici. L'idea fondamentale che è sottesa all'approccio SGR è rappresentata da quella che possiamo identificare come la trasformazione dell'informazione prodotta da un processo di faking reso attivo a seguito di una specifica richiesta, detta genericamente *query* (es. una domanda o item di un questionario di autovalutazione). Nell'accezione SGR, il faking è inteso come una procedura probabilistica che permette di trasformare una informazione (d) interna al sistema cognitivo di un

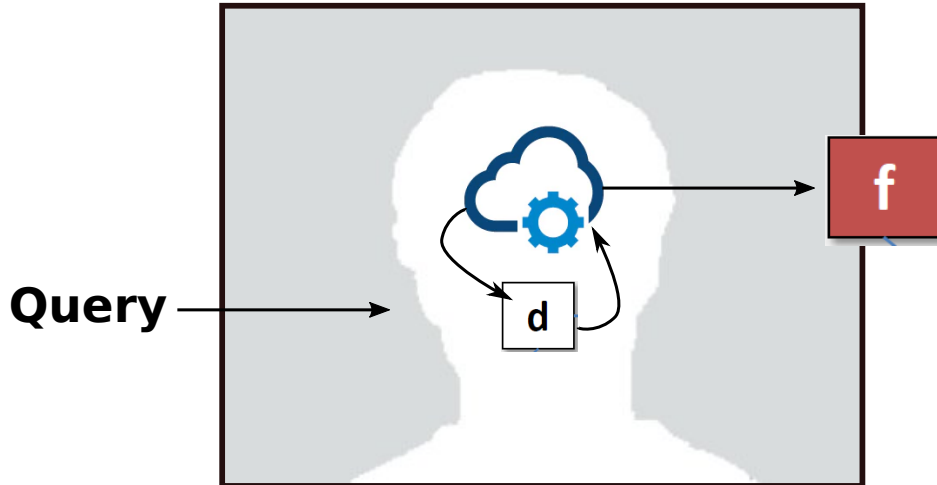


Figura 1.1: Rappresentazione dello schema SGR: l'individuo sottoposto ad uno stimolo, in una prima fase identifica la risposta onesta, d , e in una seconda elabora, e poi esprime, una risposta manipolata, f .

individuo, creatasi a seguito della richiesta esterna (query), in una nuova informazione (f), possibilmente differente dal dato interno d , che costituisce il dato osservato o la risposta data dall'individuo alla query. La Figura 1.1 illustra lo schema generale sotteso all'approccio SGR. L'assunzione di semplificazione fatta dall'approccio SGR è che la scala di informazione associata a d ed f sia la medesima (es. entrambe sono risposte di tipo dicotomico, SI-NO).

Nello schema SGR, il processo di faking è rappresentato attraverso due distinte fasi sequenziali. Nella prima fase, si cerca di modellare il processo generativo che produce la risposta originaria interna d (indicata anche come la risposta vera associata alla query). Nella seconda fase, la risposta interna d viene trasformata in una nuova risposta, f , attraverso un modello probabilistico chiamato *replacement distribution* che emula un processo di manipolazione associato ad uno specifico meccanismo di faking. Questa rappresentazione può essere facilmente generalizzata al caso di più individui e più domande (item) come viene illustrato di seguito.

Le equazioni di seguito descritte costituiranno il sistema di riferimento formale e probabilistico che verrà utilizzato per rappresentare i differenti studi sull'impatto delle manipolazioni di tipo faking sui risultati delle statistiche che saranno oggetto dei prossimi capitoli.

La matrice di dati non corrotta (informazioni interne) è indicata con \mathbf{D} e ha dimensioni $n \times m$ (n individui che rispondono a m domande di un questionario). Le righe della matrice \mathbf{D} , indicate come \mathbf{d}_i ($i = 1, \dots, n$), contengono i pattern individuali di risposte di tipo interno; le colonne della matrice \mathbf{D} , denotate come \mathbf{d}_j ($j = 1, \dots, m$), contengono tutte le risposte di tipo interno generate dai n individui per lo stesso item j .

In generale, \mathbf{D} contiene le cosiddette risposte oneste (precedenti la fase di perturbazione prodotta dal meccanismo di faking). Si assume che ogni cella d_{ij} della matrice \mathbf{D} sia l'esito di una variabile casuale discreta X_{ij}^d con valori naturali compresi in un intervallo limitato $V_q^v = \{q, q+1, \dots, v-1, v\}$ dove $q < v$. Di conseguenza, il pattern di risposte non corrotto \mathbf{d}_i è l'esito di una variabile casuale multidimensionale $(X_{i1}^d, X_{i2}^d, \dots, X_{im}^d)$ che ha probabilità $p(\mathbf{d}_i | \boldsymbol{\theta}_d)$, dove $\boldsymbol{\theta}_d$ indica il vettore di parametri che definiscono il modello non corrotto (il modello delle risposte vere). Infine, assumendo che i pattern di risposta dei vari individui siano indipendenti tra loro e pure identicamente distribuiti, si ottiene la formula generale:

$$p(\mathbf{D} | \boldsymbol{\theta}_d) = \prod_{i=1}^n p(\mathbf{d}_i | \boldsymbol{\theta}_d). \quad (1.1)$$

Il modello definito nell'equazione 1.1 viene chiamato modello vero, oppure modello non corrotto, e rappresenta la generazione dei dati senza l'intervento di alcun tipo di manipolazione.

A differenza delle risposte vere codificate nella struttura \mathbf{D} , le risposte osservabili, che sono potenzialmente corrotte dal processo di faking, sono codificate all'interno di una matrice \mathbf{F} della stessa dimensione di \mathbf{D} . Nel nostro contesto, la cella f_{ij} di \mathbf{F} rappresenta l'esito di una variabile casuale discreta X_{ij}^f che indica la risposta comunicata dall'individuo i -esimo a seguito della presentazione dell'item j . L'assunzione principale dell'approccio SGR riguarda il modo con il quale le risposte osservate f_{ij} vengono generate. In particolare, la variabile casuale discreta X_{ij}^f a cui è associata la risposta f_{ij} è assunta essere distribuita secondo una distribuzione di probabilità condizionale $p(f_{ij} | d_{ij}, \boldsymbol{\theta}_f)$ dove $\boldsymbol{\theta}_f$ indica l'insieme dei parametri del modello di sostituzione. Nella distribuzione di probabilità condizionata viene assunto che ogni risposta osservata f_{ij} dipenda esclusivamente dalla corrispondente risposta non corrotta d_{ij} e dai parametri del modello di sostituzione $\boldsymbol{\theta}_f$. In sintesi, la caratterizzazione probabilistica del profilo delle risposte osservate \mathbf{f}_i è data dall'equazione

$$p(\mathbf{f}_i | \mathbf{d}_i, \boldsymbol{\theta}_f) = \prod_{j=1}^m p(f_{ij} | d_{ij}, \boldsymbol{\theta}_f) \quad (1.2)$$

che implica un assunto di indipendenza stocastica nel processo di perturbazione attraverso gli m item. Infine, la distribuzione marginale del pattern osservato \mathbf{f}_i è dato da

$$p(\mathbf{f}_i | \boldsymbol{\theta}_d, \boldsymbol{\theta}_f) = \prod_{j=1}^m \sum_{h \in V_q^v} p(h | \boldsymbol{\theta}_d) p(f_{ij} | h, \boldsymbol{\theta}_f) \quad (1.3)$$

Uno degli obiettivi fondamentali dell'approccio SGR è stato quello di istanziare la forma analitica di queste distribuzioni di probabilità condizionali (chiamate *replacement distribution*). Un modo per costruire tali distribuzioni condizionali, che mimano il processo di perturbazione applicato ai dati originali interni (dati non corrotti), è stato descritto da Pastore e Lombardi (2014) e Lombardi et al. (2014). L'istanziamento proposta da Pastore e Lombardi (2014) fa uso di una trasformazione della funzione di densità B (beta density function).

In particolare, la distribuzione beta, $B(\gamma, \delta)$, è una funzione di densità definita nell'intervallo $[0, 1]$ ed è identificata dai parametri γ e δ , entrambi definiti in \mathbb{R}^+ . Al contrario, la distribuzione Beta generalizzata (Whitby, 1971) è una generalizzazione della beta che permette di espandere il dominio su cui è definita. Nella Beta generalizzata il dominio della variabile casuale continua è l'intervallo generico $[a, b]$, con $a, b \in \mathbb{R}$, dove a identifica il limite inferiore della variabile e b il suo limite superiore ($a < b$). Per tutti i valori all'esterno dell'intervallo $[a, b]$ la funzione di densità è pari a zero.

La funzione della beta generalizzata (G) è identificata dalla seguente equazione (dove $a \leq x \leq b$):

$$G(x; a, b, \gamma, \delta) = \frac{1}{B(\gamma, \delta)(b-a)^{\gamma+\delta-1}}(x-a)^{\gamma-1}(b-x)^{\delta-1}, \quad (1.4)$$

dove $B(\gamma, \delta)$ è la funzione beta e γ e δ sono i parametri che definiscono la forma della curva.

Sebbene in linea generale la distribuzione G si presti a modellare differenti meccanismi di faking (es. faking good a vari livelli di intensità), ciò potrebbe avvenire solo nel caso di variabili di risposta i cui valori sono assunti essere realmente continui. Per poter adattare la distribuzione G a meccanismi di perturbazione legati a variabili di tipo discreto, Pastore e Lombardi (2014) hanno proposto una procedura di discretizzazione della beta generalizzata. La procedura di discretizzazione richiede come primo passaggio di identificare una variabile casuale discreta X con dominio di valori in $R_X = \{a, a+1, \dots, b-1, b\}$ dove $a, b \in \mathbb{N} \cup \{0\}$, quindi definiscono la seguente distribuzione beta discreta generalizzata per la nuova variabile X :

$$DG(x; a, b, \gamma, \delta) = \begin{cases} \frac{G^*(x; a, b, \gamma, \delta)}{\sum_{x' \in R_X} G^*(x'; a, b, \gamma, \delta)}, & x \in R_X \\ 0, & x \notin R_X \end{cases} \quad (1.5)$$

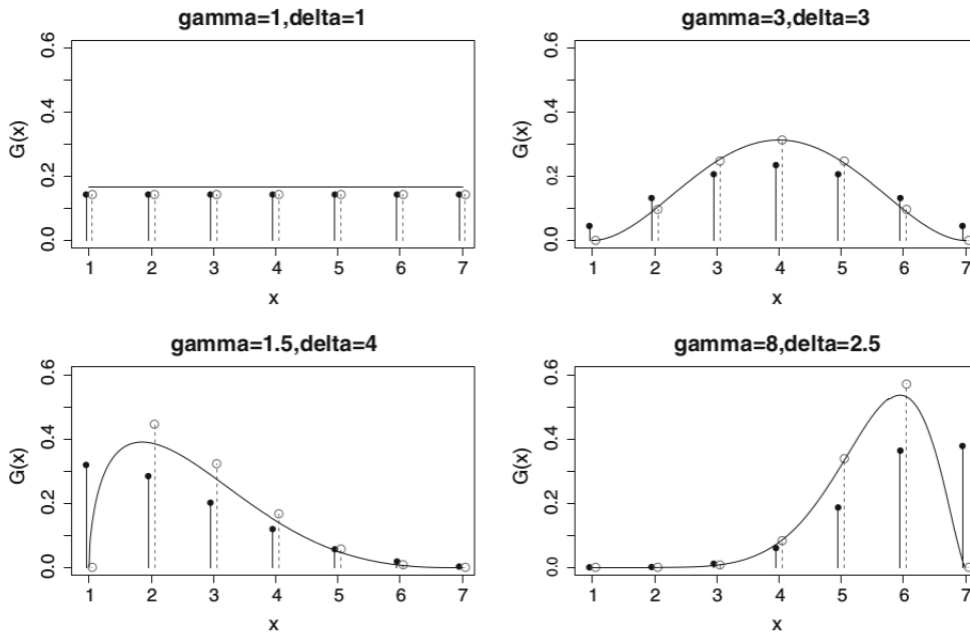


Figura 1.2: Nella Figura sono rappresentati quattro esempio di distribuzione beta G con una variabile continua definita in $[a = 1, b = 7]$. Inoltre sono rappresentati i due tipi di discretizzazione senza, con i pallini vuoti, e con correzione, pallini pieni.

dove G^* è una versione modificata della distribuzione G definita come

$$G^*(x; a, b, \gamma, \delta) = \frac{1}{B(\gamma, \delta)(b - a + 2)^{\gamma + \delta - 1}} (x - a + 1)^{\gamma - 1} (b - x + 1)^{\delta - 1}. \quad (1.6)$$

Da notare che il denominatore nell'equazione 1.5 agisce come fattore di normalizzazione per la distribuzione di massa di DG (distribuzione beta generalizzata discreta). Inoltre, nell'equazione 1.6 sono state aggiunte due costanti positive rispetto all'equazione 1.4 che funzionano come termini di correzione per la segmentazione delle code della distribuzione beta generalizzata.

Nella Figura 1.2 è illustrato un esempio di confronto tra le due distribuzioni beta discreta generalizzata (nelle versioni con e senza correzioni). La distribuzione identificata dai pallini bianchi rappresenta quella senza correzione (DG basata su G); la distribuzione identificata dai pallini neri quella con la correzione (DG basata su G^*).

Il passaggio finale consiste nell'utilizzare la nuova distribuzione DG per modellare la distribuzione di probabilità condizionata $p(f_{ij}|d_{ij}, \theta_f)$. Nello specifico, data una risposta onesta $h \in V_q^v$, la probabilità di ottenere una risposta manipolata $k \in V_q^v$ è

descritta dalla seguente equazione:

$$p(f_{ij} = k | d_{ij} = h, \boldsymbol{\theta}_f) = \begin{cases} 1, & h = k = v \\ \pi DG(k; h + 1, v, \gamma, \delta), & q \leq h < k \leq v \\ 1 - \pi, & q \leq k = h < v \\ 0, & q \leq k < h \leq v \end{cases} \quad (1.7)$$

dove π definisce la probabilità di sostituzione di h con k (dove $h \neq k$) e $1 - \pi$ rappresenta la probabilità che k sia uguale a h . Il parametro $\pi \in [0, 1]$ agisce come un peso per riscaldare la distribuzione DG , mentre $\boldsymbol{\theta}_f = (\gamma, \delta, \pi)$ costituisce il vettore di parametri del modello di faking, ed infine DG è la distribuzione beta generalizzata discreta con limiti $a = h + 1$ e $b = v$.

Nella Figura 1.3 sono riportati alcuni esempi di applicazione della distribuzione condizionale di sostituzione. Si osservi, che sebbene l'equazione 1.7 permetta di modellare meccanismi di faking in cui la persona aumenta il punteggio dell'item ($k \geq h$), processo identificato in letteratura con il nome di *faking good*, la stessa equazione può essere facilmente modificate per rappresentare il processo speculare delle perturbazioni di tipo ($k \leq h$) note come *faking bad*.

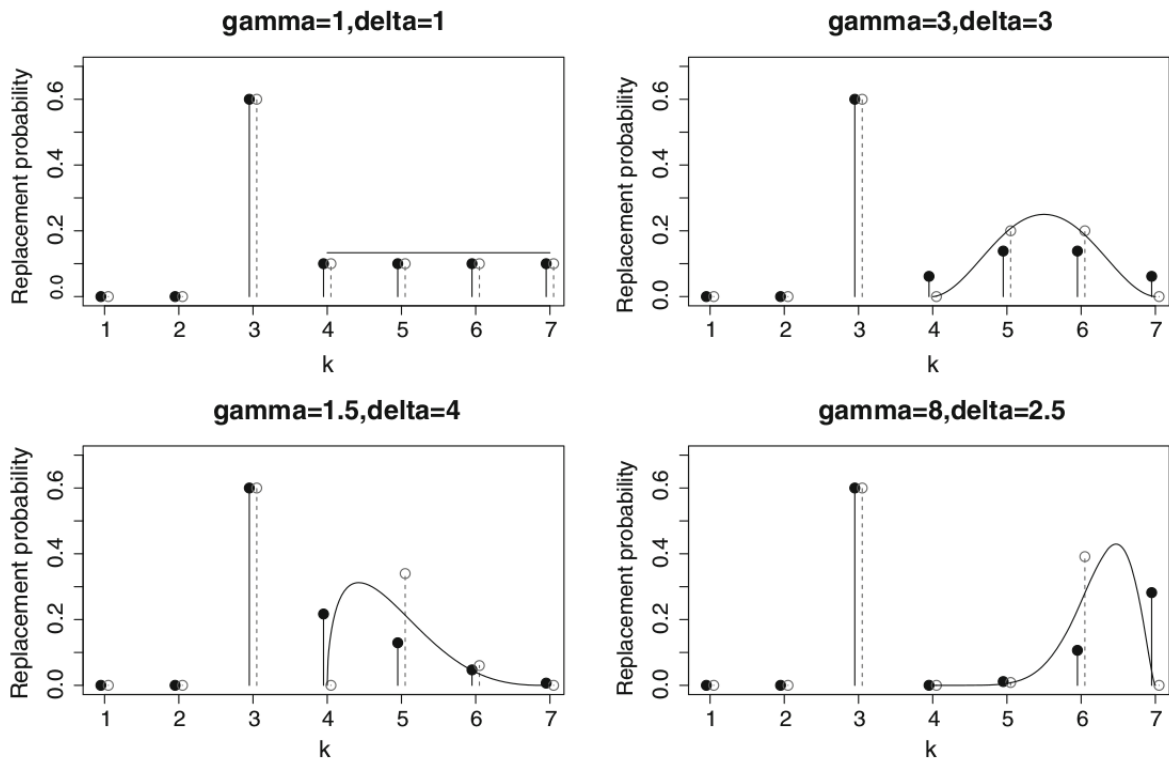


Figura 1.3: Nella Figura sono rappresentati quattro esempi di distribuzione di sostituzione con una variabile random discreta a 7 valori, valore non corretto $h = 3$ e probabilità di sostituzione $\pi = .4$. Inoltre sono rappresentati i due tipi di discretizzazione senza, con i pallini vuoti, e con correzione, pallini pieni.

Capitolo 2

L'effetto del faking sulla correlazione di due item ordinali

2.1 Introduzione

Ci sono alcuni bias ben noti che rappresentano una seria minaccia alla misura della correlazione empirica. Per esempio, la presenza dell'errore di misura nei dati non permette di osservare il vero valore di correlazione tra le variabili perché queste sono sempre affette da un errore di misurazione. Questo è uno degli assunti fondamentali della la Teoria Classica dei Test (Lord, 1980). In questo contesto, la misurazione osservata è intesa come la somma di un valore vero, ma sconosciuto, con il suo errore di misura. In generale, maggiore è l'errore di misura, meno accurate sono le stime sul campione. Un'assunzione largamente accettata della Teoria Classica dei Test è che l'errore di misura sia normalmente distribuito con media zero e varianza sconosciuta. Sotto quest'assunzione, si possono facilmente costruire intervalli di confidenza per quantificare l'incertezza delle correlazioni stimate. In ogni caso, questo semplice approccio per trattare il rumore dovuto alla misura potrebbe essere discutibile nel caso in cui si abbiano osservazioni le quali siano state corrotte da una fonte di rumore più complessa, per esempio errori asimmetrici o sistematici.

Un'altra minaccia che è collegata parzialmente alla precedente, è la presenza di outliers (osservazioni estreme) nel campione le quali possono andare ad interferire nel processo di misura sia attenuando che incrementando il valore della correlazione osservata. In questa situazione, si possono utilizzare specifiche procedure di analisi dei dati (es. approcci basati su metodi grafici o metodi statistici basati sui concetti di distanza o densità) per mitigare l'effetto degli outliers o per identificarli ed eventualmente rimuoverli dall'analisi. Inoltre, analisi di robustezza possono essere utilizzate per valutare se le osservazioni estreme possano significativamente influenzare i risultati

finali dell'analisi della correlazione (Wilcox, 2016).

Anche il trattare le variabili ordinali come se fossero continue può essere considerata una minaccia alla corretta stima della correlazione, almeno quando vengono utilizzati modelli di correlazione inadeguati. Dato che le variabili caratterizzate da un livello di misura ordinale sono largamente diffuse sia negli studi sociali che psicologici, tecniche di correlazioni più sofisticate (es., la correlazione policorica) sarebbero da preferirsi rispetto alle misure di associazione lineare di Pearson e alla correlazione non parametrica di Spearman. Purtroppo nelle maggior parte dei casi tali tecniche più sofisticate risultano essere difficilmente applicabili in quanto non sempre implementate nei principali applicativi statistici riducendone così di molto la loro diffusione. Similmente, dividere una variabile teoricamente continua in un numero finito di categorie discrete ha un effetto sulla correlazione stimata, la quale dovrebbe essere corretta o calcolata con specifici modelli di analisi (es., attraverso l'utilizzo della correlazione policorica). Questo fenomeno è conosciuto come "broad categories" o "few ordinal categories" ed è stato ampiamente studiato nella letteratura psicometrica (Pearson, 1913).

Inoltre, in alcuni contesti particolari, ad esempio nella psicologia dell'educazione o in psicologia del lavoro nei contesti di selezione del personale, i ricercatori hanno solitamente accesso solo ai dati di una popolazione ristretta (*range restriction bias*) sebbene siano interessati a stimare i parametri per la popolazione senza restrizioni. Alcune classiche soluzioni sono state proposte nella letteratura statistica per mitigare questo bias con specifiche formule di correzione (p.e., Spearman's correction of attenuation, Kelley's correction for range restriction, Spearman, 1904; Kelley, 1947).

Infine, in alcuni questionari psicologici, la necessità di analizzare dati empirici fa spesso emergere il problema di possibili osservazioni menzognere o fraudolente nei dati. Chiamiamo questo problema la *minaccia dei dati fake* (fake data bias). Questo aspetto è particolarmente rilevante per quelle ricerche che indagano su tematiche sensibili come, per esempio, comportamenti sessuali a rischio o consumo di droga oppure in contesti clinici dove l'individuo tende a esagerare i sintomi (malingering) in una check list per diagnosi clinica allo scopo di raggiungere determinati obiettivi come, per esempio, evitare di essere accusato di un crimine (p.e., Mittenberg, Patton, Canyock, & Condit, 2002; Hall & Hall, 2007). Il problema principale è che spesso per alcuni tipi di misure self-report non ci sono in realtà le basi per assumere che i partecipanti stiano rispondendo onestamente, non c'è una strada semplice per verificare le risposte e nemmeno una metodologia robusta per identificare la presenza di risposte fake nei dati osservati (Lombardi & Pastore, 2015). Come per la minaccia degli outlier, la presenza nel campione di osservazioni manipolate da fake può artificialmente incrementare, o diminuire, l'associazione tra due variabili (Ellingson et al., 1999; Zickar & Robie, 1999;

Pauls & Crost, 2005; Ziegler & Buehner, 2009; Galić et al., 2012).

In questo capitolo, ci concentreremo sull'ultimo tipo di bias ed esploreremo l'impatto che dati fake hanno nell'interpretazione dei risultati statistici calcolati utilizzando i più noti e diffusi modelli statistici per la stima della correlazione tra due variabili ordinali. In particolare, la minaccia dei dati fake implica una domanda cruciale: se i dati che osserviamo includessero risposte fake, fino a che punto la correlazione che osserviamo potrebbe essere differente da quella reale? In altre parole, quale percentuale di soggetti fakers (all'interno di un campione) e che tipo di processo di manipolazione è avvenuta in modo da modificare i risultati (es., stima della correlazione, valutazione del fit del modello) da essere in qualche modo differenti da quelli reali? Anche nella situazione più semplice (risposte fake casuali), la risposta non è necessariamente ovvia in quanto perfino la perturbazione random dei dati tende a diminuire l'accuratezza delle inferenze statistiche.

Oggigiorno vi è una vasta letteratura riguardo il problema del faking che spazia su diversi aspetti del problema, orientati sia all'analisi dei dati che alla definizione di costrutti identificativo il processo di faking. Per esempio, la problematica del faking è stata investigata utilizzando paradigmi empirici ad hoc come ad lib faking o addestramento al faking per raccogliere dati e simulare pattern manipolati da faking (Zickar & Robie, 1999; Zickar, Gibby, & Robie, 2004). In particolare, recentemente alcuni autori hanno proposto un metodo razionale per valutare i dati fake nel contesto della desiderabilità sociale o situazioni di motivazione al faking utilizzando approcci di analisi fattoriale (Ferrando, 2005; Ferrando & Anguiano-Carrasco, 2009, 2011) così come factor mixture models (Leite & Cooper, 2010). Purtroppo, minore attenzione è stata indirizzata ad aspetti più generali collegati all'impatto del faking nei risultati statistici. Per esempio, in che modo i dati e i risultati delle analisi statistiche possono essere sensibili alle manipolazioni conseguenti al processo di faking della risposta? In questo contributo adotteremo un approccio statistico, chiamato Sample Generation by Replacement (SGR, Lombardi & Pastore, 2012) per analizzare l'incertezza delle inferenze dovute a possibili dati fake così come per valutare le implicazioni del faking sui i risultati di alcuni semplici modelli statistici.

In particolare, in questo capitolo proponiamo un'innovativa prospettiva analitica per la rappresentazione SGR, la quale permette di studiare l'effetto del faking sulla correlazione tra due variabili ordinali al livello di popolazione (abbiamo chiamato questa estensione del paradigma mixture SGR, M-SGR). Inoltre, valuteremo la bontà di questa rappresentazione della popolazione comparando il calcolo ottenuto dal modello teorico della popolazione con la correlazione stimata attraverso la mediana di correlazioni empiriche generate secondo un classico schema di tipo Monte Carlo. Quest'ultima

procedura sarà inoltre utile per valutare l'impatto di diverse dimensioni campionarie e la loro deviazione dalla stima della popolazione. In particolare, il problema della minaccia dei dati fake è testato usando due indici di correlazione molto diffusi (correlazione di Pearson e correlazione di Spearman), calcolati su una coppia di variabili ordinali.

Usando il paradigma SGR mostreremo come, sotto alcune condizioni di faking, sia sempre possibile sia incrementare che diminuire la correlazione tra due variabili ordinali in un modo predicibile e sistematico. Il risultato più interessante e al tempo stesso allarmante è che una correlazione statisticamente significativa, ottenuta attraverso un questionario self report, potrebbe essere semplicemente l'effetto di una falsa associazione positiva dovuta a una strategica manipolazione dei soggetti. Simmetricamente, alcuni risultati significativi potrebbero essere stati nascosti nella correlazione empirica perché le osservazioni sono state perturbate dal processo di faking.

2.2 Due misure di correlazione

In questo studio sono state esaminate due misure di correlazione: il coefficiente di correlazione lineare (correlazione di Pearson) e il coefficiente di correlazione a ranghi di Spearman. Sebbene altri tipi di indici di correlazione potrebbero essere utilizzati nella valutazione della forza dell'associazione tra due variabili ordinali (es., Tau-b di Kendall, s di Somers, la statistica Gamma), in questo studio ci siamo limitati a valutare l'impatto delle manipolazioni di tipo fake per le due misure di associazione più largamente diffuse in letteratura: la correlazione lineare di Pearson e la correlazione non parametrica di Spearman. Nel nostro contesto, la versione discreta della correlazione di Pearson è utilizzata come modello di riferimento, seppur teoricamente improprio, per il confronto con un modello più adeguato, la correlazione di Spearman, una misura non parametrica di associazione tra due variabili ordinali. Di seguito introduciamo i termini principali e la notazione utile alla descrizione dei due modelli di correlazione nel contesto di variabili casuali discrete con range finito.

Definiamo X e Y due variabili casuali discrete con range finito comune $\{1, 2, \dots, v\} \subset \mathbb{R}$ con $v \in \mathbb{N}$. Inoltre, sia p_{ij} la notazione per descrivere la distribuzione di probabilità congiunta $P(X = i, Y = j)$ (con $i, j = 1, \dots, v$), e siano p_i e p_j le corrispondenti probabilità marginali. Le distribuzioni cumulative marginali sono calcolate come $F_i = \sum_{k=1}^i p_k$ e $G_j = \sum_{k=1}^j p_k$ per X e Y , rispettivamente. Infine, sia $(\mathbf{x}, \mathbf{y}) = ((x_1, y_1), \dots, (x_n, y_n))$ il campione bivariato contenente n distinte osservazioni estratte dalla distribuzione discreta della popolazione. Siamo adesso in grado di descrivere dettagliatamente le due misure di correlazione in oggetto.

2.2.1 Indice di correlazione di Pearson

La correlazione parametrica di Pearson fu sviluppata per misurare la dipendenza lineare tra due variabili continue, anche se nella pratica empirica applicata è norma diffusa usarla anche per variabili numeriche di tipo discrete (es., item di tipo Likert) forzando la naturale scala ordinale della misura originale. In un contesto ideale, la relazione tra le due variabili è lineare e le deviazioni dalla linea retta generalmente attenuano la magnitudine del coefficiente di correlazione. In aggiunta, le due variabili sono assunte distribuite normalmente e dovrebbero soddisfare il principio di omoschedasticità (la varianza di una variabile è costante sul proprio range di valori ed è indipendente dai valori che assume l'altra variabile). Sfortunatamente, nelle variabili discrete, l'applicazione della correlazione di Pearson non rispetta alcuno di questi assunti. Per tali ragioni l'applicazione di questo tipo di correlazione parametrica è nel caso migliore sub-ottimale.

Correlazione campionaria

La versione discreta della correlazione di Pearson, indicata da r_p , può essere calcolata usando la covarianza campionaria, $\hat{\sigma}_{xy}$, e le varianze campionarie, $\hat{\sigma}_x$ e $\hat{\sigma}_y$. In particolare, la formula di r_p è definita:

$$r_p = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y} = \frac{\sum_{s=1}^n (x_s - \bar{x})(y_s - \bar{y})}{\sqrt{\sum_{s=1}^n (x_s - \bar{x})^2} \sqrt{\sum_{s=1}^n (y_s - \bar{y})^2}} \quad (2.1)$$

dove \bar{x} e \bar{y} indicano, rispettivamente, le due medie campionarie per i campioni discreti \mathbf{x} e \mathbf{y} .

Correlazione della popolazione

Al livello di popolazione, il coefficiente di Pearson è usualmente denotato dalla lettera greca ρ_p . La formula per la correlazione a livello di popolazione per variabili discrete è la seguente:

$$\rho_p = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{\left[\sum_{i=1}^v \sum_{j=1}^v ij p_{ij} \right] - \left[\sum_{i=1}^v i p_{i.} \right] \left[\sum_{j=1}^v j p_{.j} \right]}{\sqrt{\sum_{i=1}^v i^2 p_{i.} - \left[\sum_{i=1}^v i p_{i.} \right]^2} \cdot \sqrt{\sum_{j=1}^v j^2 p_{.j} - \left[\sum_{j=1}^v j p_{.j} \right]^2}} \quad (2.2)$$

dove σ_{xy} , σ_x , σ_y , sono, rispettivamente, le covarianze e varianze della popolazione e n è la numerosità dei campionaria.

2.2.2 Indice di correlazione di Spearman

Il coefficiente di correlazione di Spearman è una misura non parametrica di associazione tra due variabili; è basato sui ranghi e rappresenta una delle prime misure di correlazione ad essere stata presentata nella letteratura statistica. Necessita che entrambe le variabili siano misurate almeno su scala ordinale in maniera che le osservazioni di \mathbf{x} e \mathbf{y} possano essere classificate in due insiemi ordinati. L'assunto principale della correlazione di Spearman è che le due variabili siano monotonicamente relate l'una l'altra.

Correlazione campionaria

A livello campionario, la correlazione di Spearman è generalmente descritta dalla seguente formula:

$$r = \frac{\hat{\sigma}_{rank(\mathbf{x}, \mathbf{y})}}{\hat{\sigma}_{rank(\mathbf{x})} \hat{\sigma}_{rank(\mathbf{y})}} = \frac{\sum_{s=1}^n (R_s - \bar{R})(S_s - \bar{S})}{\sqrt{\sum_{s=1}^n (R_s - \bar{R})^2 \cdot \sum_{s=1}^n (S_s - \bar{S})^2}} \quad (2.3)$$

dove R_i e S_i sono i ranghi, rispettivamente, delle osservazioni x_i e y_i e \bar{R} e \bar{S} le due medie campionarie.

Correlazione della popolazione

La versione a livello di popolazione della correlazione di Spearman per variabili discrete a supporto finito è stata descritta in letteratura solo recentemente (si veda Nešlehová, 2007). La formula della popolazione proposta per variabili discrete è definita come

$$\rho = \frac{3 \sum_{i=1}^v \sum_{j=1}^v p_{ij} [(F_i + F_{i-1})(G_j + G_{j-1}) - 1]}{\sqrt{(1 - \sum_{i=1}^v p_{i.}^3)(1 - \sum_{j=1}^v p_{.j}^3)}} \quad (2.4)$$

La versione campionaria della formula 2.4 può essere facilmente ottenuta rimpiazzando i termini della popolazione $p_{ij}, p_{i.}, p_{.j}, F_i$ e G_j con i corrispettive stime campionarie calcolate utilizzando la tabella di contingenza di ordine $v \times v$ del campione bivariato (\mathbf{x}, \mathbf{y}) . Può essere dimostrato (si veda Nešlehová, 2007) che la versione campionaria della formula 2.4 si riduce alla ben nota formula campionaria 2.3 quando X e Y sono variabili casuali discrete definite in un intervallo finito.

2.3 SGR tradizionale

La metodologia SGR è una modellizzazione atta a simulare il processo di faking. Essa è caratterizzata da due stadi e utilizza due distinti modelli. Nel primo stadio si generano i dati "onesti" secondo uno specifico modello generativo. In questa prima fase, i dati così generati riflettono come idealmente dovrebbe essere il campione senza osservazioni corrotte dal faking. Il secondo stadio è un processo di sostituzione dove, attraverso un modello di perturbazione, i dati vengono perturbati in modo da riflettere un flusso di informazioni corrotto. Nell'SGR standard, la prima procedura è realizzata attraverso delle semplici tecniche Monte Carlo (MC), mentre la seconda procedura è modellata adottando modelli probabilistici ad-hoc (es., Lombardi & Pastore, 2012; Lombardi et al., 2014; Pastore & Lombardi, 2014). Negli ultimi anni l'impianto SGR è stato prevalentemente utilizzato per lo studio della sensibilità di indici di fit su dati manipolati da faking secondo simulazioni di tipo MC. A differenza degli studi precedenti basati sull'approccio SGR, in questo capitolo presteremo più attenzione ad alcuni aspetti dell'effetto del faking a livello della popolazione degli indici statistici di associazione. In particolare, verrà analizzato l'indice di Cohen (denotato q nella sua versione campionaria e Q in quella della popolazione), un indice di dimensione dell'effetto per le differenze di correlazioni. Inoltre, per meglio evidenziare la natura a livello di popolazione del problema del faking, verrà lievemente modificata la notazione dell'SGR standard introducendo una nuova rappresentazione matriciale. Definiamo (X^d, Y^d, X^f, Y^f) un insieme ordinato di quattro variabili casuali discrete, definite nello stesso dominio di valori $\{1, 2, \dots, v\}$. Nella rappresentazione SGR, le quattro variabili possono essere distinte in due gruppi, definendo le condizioni onesta/non corrotta, $\{X^d, Y^d\}$, e la condizione di faking, $\{X^f, Y^f\}$. La distribuzione di probabilità congiunta per la condizione onesta è rappresentata dalla seguente formula:

$$p_{ij}^d = P(X^d = i, Y^d = j | \boldsymbol{\theta}_d) \quad (2.5)$$

con $(i, j) \in \{1, 2, \dots, v\}^2$ e dove $\boldsymbol{\theta}_d$ è l'array di parametri associato al modello non corrotto. Per contrasto, la condizione di faking è rappresentata attraverso una distribuzione di probabilità condizionata:

$$z_{hk|ij} = P(X^f = h, Y^f = k | X^d = i, Y^d = j, \boldsymbol{\theta}_f) \quad (2.6)$$

con $(i, j, h, k) \in \{1, 2, \dots, v\}^4$ e dove $\boldsymbol{\theta}_f$ è l'array di parametri associati al modello di faking. L'equazione 2.6 identifica la cosiddetta *replacement distribution*, distribuzione di sostituzione, in un modello SGR. Questa distribuzione rappresenta la probabili-

tà condizionale di sostituzione della risposta originariamente osservata, (i, j) , sotto il modello non corrotto, con il nuovo valore manipolato, (h, k) , e costituisce il nucleo centrale di ogni rappresentazione SGR (Lombardi & Pastore, 2012). La distribuzione di probabilità congiunta nella condizione di faking è quindi la probabilità marginale

$$z_{hk} = \sum_{i=1}^v \sum_{j=1}^v p_{ij}^d z_{hk|ij} \quad (2.7)$$

Un ulteriore assunto semplificativo nel contesto SGR richiede che la distribuzione di sostituzione soddisfi la proprietà di *indipendenza condizionata*:

$$z_{hk|ij} = z_{h|i}^1 z_{k|j}^2 \quad (2.8)$$

dove $z_{h|i}^1$ e $z_{k|j}^2$ sono due distribuzioni condizionate separate, rispettivamente, $P(X^f = h|X^d = i, \boldsymbol{\theta}_{f_X})$ e $P(X^f = k|X^d = j, \boldsymbol{\theta}_{f_Y})$. Quindi l'equazione 2.7 si riduce a

$$z_{hk} = \sum_{i=1}^v \sum_{j=1}^v p_{ij}^d z_{h|i}^1 z_{k|j}^2. \quad (2.9)$$

Da notare che le due distribuzioni condizionate sono caratterizzate da differenti parametrizzazioni, $\boldsymbol{\theta}_{f_X}$ e $\boldsymbol{\theta}_{f_Y}$. Questo riflette l'idea che i valori sostituiti seguono esclusivamente il modello originale (non corrotto) corrispondente e lo specifico processo di faking. Inoltre il processo di faking potrebbe essere differente per le due variabili X e Y sulla base dei valori dei parametri $\boldsymbol{\theta}_{f_X}$ e $\boldsymbol{\theta}_{f_Y}$. Per esempio, potremmo decidere di adottare due differenti direzioni e intensità di faking quando un rispondente usa differenti strategie di faking per i due item/variabili X e Y .

La distribuzione marginale della componente di faking (equazione 2.9) può essere descritta in una forma compatta usando la seguente notazione matriciale:

$$\mathbf{Z} = (\mathbf{\Pi}_d^T \mathbf{Z}_1)^T \mathbf{Z}_2 \quad (2.10)$$

dove $\mathbf{\Pi}_d = [p_{ij}^d]$ è la matrice di ordine $v \times v$ che rappresenta la distribuzione congiunta per la condizione onesta, e $\mathbf{Z}_1^T = [z_{h|i}^1]$ e $\mathbf{Z}_2^T = [z_{k|j}^2]$ sono, rispettivamente, le due matrici trasposte di ordine $v \times v$ associate con le distribuzioni di sostituzione. L'equazione 2.10 può essere usata per campionare osservazioni bivariate le quali sono in linea con il modello di faking parametrizzato secondo $(\boldsymbol{\theta}_d, \boldsymbol{\theta}_{f_X}, \boldsymbol{\theta}_{f_Y})$.

2.4 Mixture SGR

In alcune circostanze empiriche, assumere da modello che tutti i partecipanti abbiano la stessa probabilità di manipolare una risposta è semplicemente irrealistico. In generale, sarebbe più utile ipotizzare che, mentre alcuni individui tendono a manipolare le risposte, altri potrebbero semplicemente rispondere sinceramente. A questo scopo, la rappresentazione SGR può essere facilmente estesa per permettere ad ogni rispondente di essere parte di uno dei due distinti gruppi: a) il gruppo degli onesti, b) il gruppo dei fakers. Se un partecipante appartiene al primo gruppo, le corrispondenti risposte verranno campionate seguendo il modello $\mathbf{\Pi}_d$. Invece, se un partecipante appartiene al gruppo dei fakers, le risposte verranno campionate sulla base del modello \mathbf{Z} . In termini più formali, la rappresentazione dell' SGR mixture prende la seguente forma:

$$s_{xy} = (1 - \alpha)p_{xy} + \alpha z_{xy}, \quad (x, y) \in \{1, 2, \dots, v\}^2 \quad (2.11)$$

dove s_{xy} è la probabilità di osservare una coppia di variabili $X = x$ e $Y = y$. Nell'equazione 2.11, il parametro $\alpha \in [0, 1]$ definisce il peso delle probabilità nel modello mixture e rappresenta la proporzione di fakers nella popolazione. In notazione matriciale l'equazione 2.11 assume la seguente forma:

$$\text{vec}(\mathbf{\Pi}_m) = (1 - \alpha)\text{vec}(\mathbf{\Pi}_d) + \alpha \mathbf{Z}_2^T \otimes \mathbf{Z}_1^T \text{vec}(\mathbf{\Pi}_d) \quad (2.12)$$

dove $\text{vec}(\cdot)$ e \otimes sono, rispettivamente, l'operatore di vettorizzazione e il prodotto di Kronecker. Da notare i due casi speciali che si ottengono quando $\alpha = 0$ oppure $\alpha = 1$. Nel primo caso la probabilità congiunta $\mathbf{\Pi}_m$ si riduce al modello onesto $\mathbf{\Pi}_d$, nel secondo si riduce al modello faking puro \mathbf{Z} .

2.4.1 Modello di risposta onesta

In generale, diverse opzioni sono disponibili per modellare la distribuzione congiunta $\mathbf{\Pi}_d$ associata al modello onesto (es., Samejima, 1969; Muthén, 1984; Moustaki & Knott, 2000; Jöreskog & Sörbom, 2001). Nel contesto SGR una scelta naturale è l'adozione di una rappresentazione multivariata di variabili latenti chiamata UVA, Underlying Variable Approach, (Muthén, 1984; Jöreskog & Sörbom, 2001). L'idea di base di questo approccio è che le due variabili discrete/ordinali osservate siano trattate come se fossero di fatto la discretizzazione di variabili numeriche latenti distribuite normalmente. Nel nostro contesto bivariato indichiamo questa coppia di variabili latenti normali con (W_1, W_2) . Nel contesto UVA la parametrizzazione $\boldsymbol{\theta}_d$ della distribuzione congiunta $\mathbf{\Pi}_d$ è un vettore contenente un set di $v - 1$ soglie le quali sono usate per discretizzare le

due variabili continue sottostanti, e un parametro di correlazione ρ_d il quale modula la loro relazione lineare. Più precisamente la parametrizzazione UVA è data da

$$p_{ij}^d = \int_{\xi_{i-1}}^{\xi_i} \int_{\xi_{j-1}}^{\xi_j} \phi(w_1, w_2 | \mathbf{0}, \mathbf{R}_d) dw_1 dw_2, \quad i = 1, \dots, v; j = 1, \dots, v \quad (2.13)$$

dove ξ_1, \dots, ξ_{v-1} sono le $v - 1$ soglie (con $\xi_0 = -\infty, \xi_v = +\infty$) e ϕ è una distribuzione standardizzata bivariata con media $0 = (0, 0)$ e la matrice di correlazione \mathbf{R}_d con il parametro di correlazione ρ_d .

2.4.2 Modelli di manipolazione fake

Oggi giorno vi è un diffuso consenso sul fatto che il faking sia una manipolazione intenzionale della risposta finalizzato a raggiungere un guadagno personale (p.e., MacCann et al., 2012). In questo capitolo limitiamo l'ambito a due scenari rilevanti di faking: a) faking good b) faking bad. Il faking good può essere definito come un tentativo cosciente di presentare informazioni false per creare una impressione più positiva con l'obiettivo di influenzare altre persone (p.e., Furnham, 1986; Zickar & Robie, 1999; McFarland & Ryan, 2000). In generale, i manipolatori fake-good modificano i loro punteggi sulla scala in modo da fornire punteggi più estremi (p.e., Furnham, 1986; Viswesvaran & Ones, 1999; McFarland & Ryan, 2000; Griffin, Hesketh, & Grayson, 2004). Nel contesto SGR, una manipolazione fake-good rappresenta una situazione dove le risposte sono soggette esclusivamente a uno spostamento positivo:

$$X^f > X^d \quad (\text{risp. } Y^f > Y^d)$$

Per esempio, una manipolazione faking good potrebbe essere associata a recensioni acquistate da un negozio online per aumentare il proprio ranking. Dall'altra parte, il faking bad indica il tentativo cosciente di creare una impressione più negativa di quella effettivamente reale fornendo così risposte con valori più bassi:

$$X^f < X^d \quad (\text{risp. } Y^f < Y^d).$$

Per esempio, in una selezione per la leva militare obbligatoria, un candidato potrebbe provare a manipolare un questionario di personalità mimando alcuni sintomi di malattia mentale con l'obiettivo di evitare la leva. Nell'approccio SGR il meccanismo di faking è descritto nei modelli di sostituzione (replacement model) \mathbf{Z}_1 e \mathbf{Z}_2 , i quali sono parametrizzati secondo una versione discreta della distribuzione beta generalizzata (descritta nel primo capitolo), i quali parametri sono, rispettivamente, $\boldsymbol{\theta}_{f_X} = (\gamma_X, \delta_X)$ e

$\theta_{f_Y} = (\gamma_Y, \delta_Y)$ (vedasi Lombardi & Pastore, 2012; Lombardi et al., 2014). In generale, questa parametrizzazione del modello è molto flessibile e può facilmente caratterizzare le due tipologie di faking con diversi livelli di intensità. In particolare, se noi settiamo i valori dei parametri come $1 \leq \gamma < \delta \leq 5$, possiamo ricostruire distribuzioni di sostituzione che vanno ad imitare uno spostamento moderato del valore originale osservato nella risposta (Figura 2.1, prima colonna). Questa configurazione può essere applicata ogniqualvolta si crede che le osservazioni originali siano state corrotte da un processo di faking good slight (Zickar & Robie, 1999; Zickar et al., 2004). Al contrario, la condizione $1 \leq \delta < \gamma \leq 5$ descrive uno scenario di faking nel quale i valori manipolati corrispondono a un grande slittamento dal valore originale (modello extreme, Figura 2.1 seconda colonna). Il modello extreme può essere adottato se si crede che le osservazioni originali siano state corrotte da un processo di faking extreme good (Zickar & Robie, 1999; Zickar et al., 2004). Da notare che è possibile ottenere facilmente una rappresentazione totalmente simmetrica la quale può essere utilizzata per modellare la condizione speculare di faking bad, ma che non sarà discussa ulteriormente qui (vedasi Figura 2.1, terza e quarta colonne). Per ulteriori dettagli, il lettore può fare riferimento ai lavori originali riguardo il modello di parametrizzazione nell'impianto SGR. (es., Lombardi & Pastore, 2012; Lombardi et al., 2014).

2.5 Valutazione della correlazione corrotta da faking a livello di popolazione

2.5.1 L'indice Q di Cohen

Con l'obiettivo di valutare l'impatto del faking nella correlazione a livello di popolazione, abbiamo adottato l'indice di dimensione dell'effetto (effect size) Q di Cohen (J. Cohen, 1988). A livello di popolazione, l'indice Q è definito come la differenza tra due trasformazioni di Fisher delle correlazioni calcolate a livello di popolazione (ρ_m) sulla base della distribuzione mista congiunta $\mathbf{\Pi}_m$ e la correlazione calcolata a livello di popolazione (ρ_d) sulla base della distribuzione congiunta originale non corrotta $\mathbf{\Pi}_d$:

$$Q = \frac{1}{2} \log \frac{1 + \rho_m}{1 - \rho_m} - \frac{1}{2} \log \frac{1 + \rho_d}{1 - \rho_d} \quad (2.14)$$

Abbiamo calcolato la statistica Q per i due indici di correlazione (Pearson e Spearman) separatamente, usando le corrispettive formule di calcolo a livello di popolazione (equazioni 2.2 e 2.4)

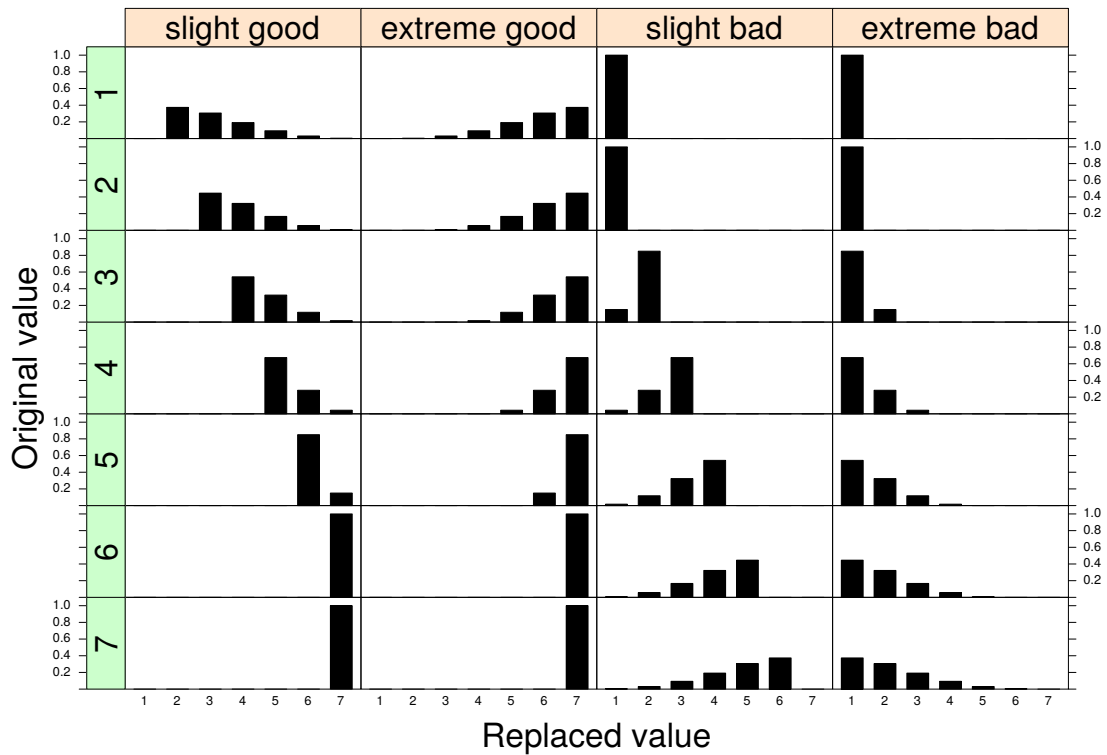


Figura 2.1: Quattro esempio di distribuzione di sostituzione per una variabile discreta a 7 livelli. Ogni colonna del grafico corrisponde a una differente distribuzione di sostituzione condizionata. Ogni riga del grafico corrisponde a uno dei 7 livelli discreti non manipolati. I parametri assegnati ai quattro modelli sono: modello *slight* ($\gamma = 1.5, \delta = 4$) e modello *extreme*: ($\gamma = 4, \delta = 1.5$). Le distribuzioni di sostituzione sono utilizzate nelle due configurazioni, rispettivamente, *faking good* e *faking bad*

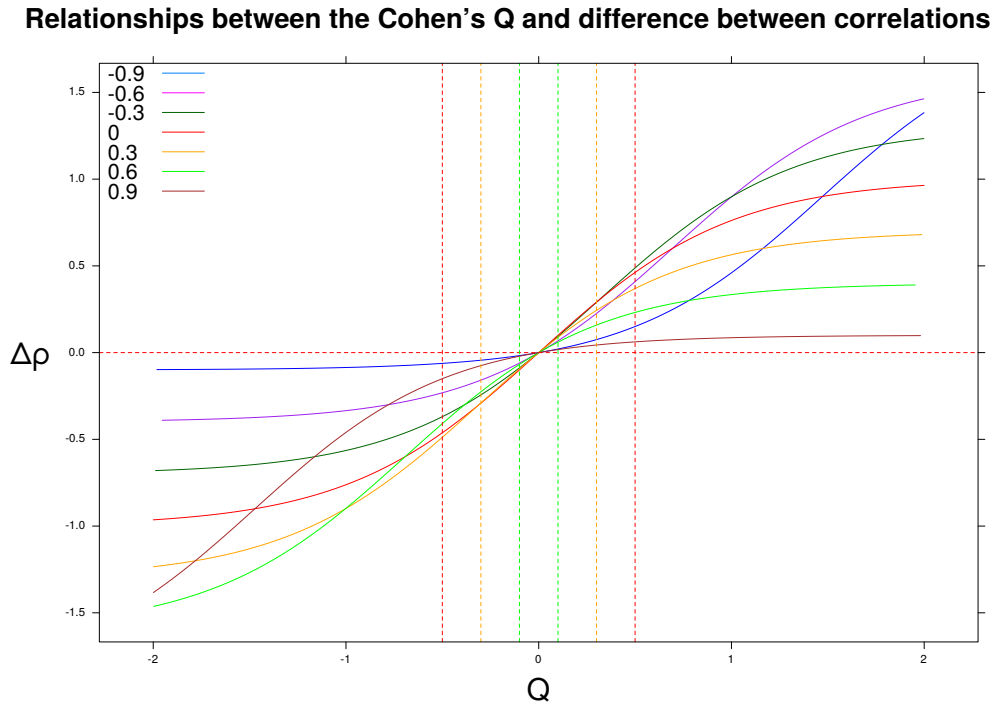


Figura 2.2: Differenza di correlazione $\Delta\rho = \rho_m - \rho_d$ come funzione della statistica Q di Cohen della correlazione originaria $\rho_d \in \{-0.9, -0.6, -0.3, 0, 0.3, 0.6, 0.9\}$. Le linee verticali identificano dimensioni degli effetti piccole, medie o grandi.

Per spiegare il significato di Q , prima di tutto definiamo $\Delta\rho = \rho_m - \rho_d$ cioè a dire il differenziale tra le due correlazioni nel contesto perturbato ed in quello onesto. La Figura 2.2 mostra l'associazione tra Q e i differenziali di correlazione $\Delta\rho$. J. Cohen (1988) fornisce una sostanziale interpretazione per differenti range di dimensione dell'effetto. Una dimensione dell'effetto *piccola*, $Q = .10$, corrisponde a piccole differenze tra la correlazione non corrotta e la correlazione della popolazione mista (es., $\rho_d = .20, \rho_m = .29$). Una dimensione dell'effetto *media*, $Q = .30$, è collegata a differenze leggere, (es., $\rho_d = .20, \rho_m = .46$). Infine, una *grande* dimensione dell'effetto, $Q = .50$, corrisponde a più ampie differenze in correlazione, (es., $\rho_d = .20, \rho_m = .61$).

2.6 Risultati dell'analisi a livello di popolazione

Come ci si attendeva, più grandi dimensioni dell'effetto sono associate a modelli di *extreme faking* e maggiori proporzioni di fakers nella popolazione mista. In generale, l'indice Q di Cohen può assumere sia valori positivi che negativi a seconda dello specifico modello di faking e della proporzione di fakers nella popolazione mista. Per esempio, sotto un modello slight e il 50% di fakers nella popolazione mista, una correlazione di Spearman non corrotta, $\rho_d = -0.60$, è affetta da un bias di $Q = 0.54$. Similmente,

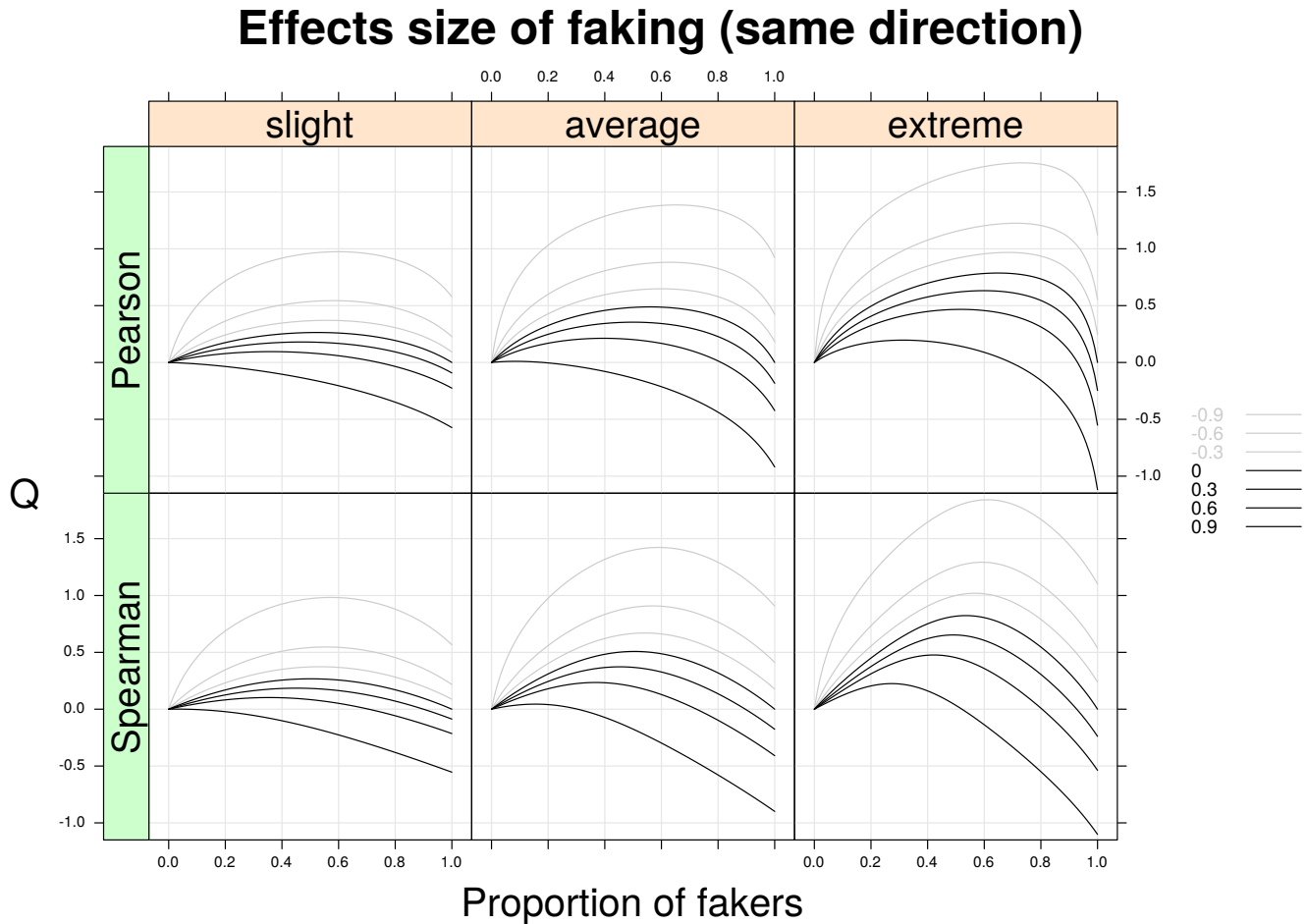


Figura 2.3: La statistica Q di Cohen come funzione della proporzione di fakers nella popolazione, modello di faking (slight, average, extreme) e della correlazione non corretta. L'analisi a livello di popolazione riguarda i due indici di correlazione (Pearson e Spearman) con lo stesso modello e direzione di faking applicato su due variabili discrete con un uguale e limitato numero di possibili livelli ($v = 5$) e distribuzioni marginali con le seguenti masse di probabilità 0.06, 0.25, 0.38, 0.25, e 0.06. (Il modello average, $\gamma = 3, \delta = 3$, rappresenta un compromesso tra il modello slight e quello extreme)

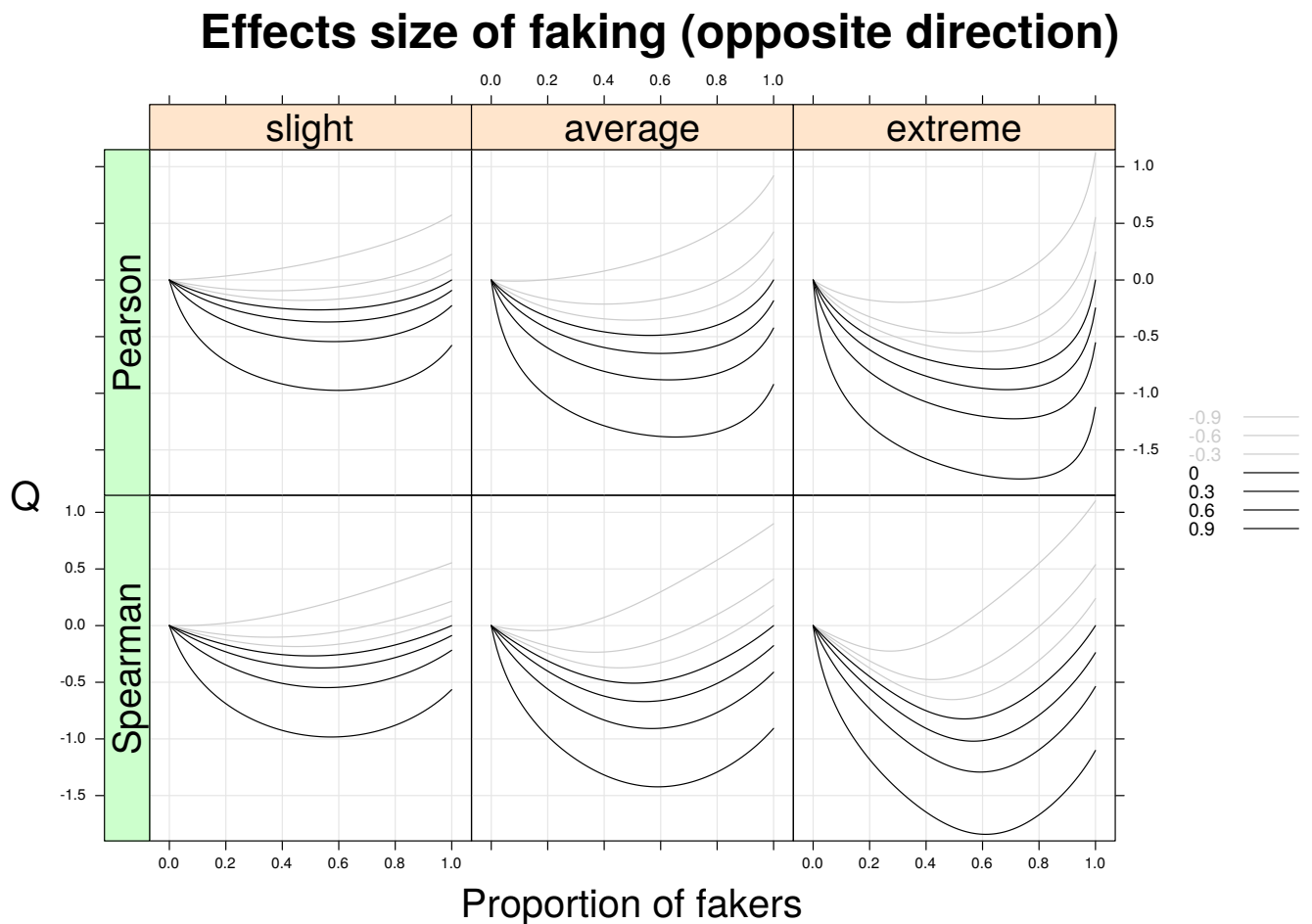


Figura 2.4: La statistica Q di Cohen come funzione della proporzione di fakers nella popolazione, modello di faking (slight, average, extreme) e della correlazione non corrotta. L'analisi a livello di popolazione riguarda i due indici di correlazione (Pearson e Spearman) sotto lo stesso modello e opposta direzione di faking applicato su due variabili discrete con un uguale e limitato numero di possibili livelli ($v = 5$) e distribuzioni marginali con le seguenti masse di probabilità 0.06, 0.25, 0.38, 0.25, e 0.06. (Il modello average, $\gamma = 3, \delta = 3$, rappresenta un compromesso tra il modello slight e quello extreme)

sotto un modello di faking extreme con il 20% di fakers nella popolazione mista, una correlazione di Spearman non corrotta pari a zero mostra un bias di $Q = 0.45$. Inoltre, se consideriamo una correlazione positiva molto alta, $\rho_d = +0.90$, con il 20% di fakers nella popolazione mista risulta in un bias di $Q = 0.2$. Al contrario, considerando per la stessa correlazione una diversa percentuale di fakers, 70%, il bias diventa negativo: $Q = -0.33$. Anche per proporzioni di fakers più piccole (es., 10%), l'effetto del faking può avere un impatto rilevante sulla correlazione originale ρ_d (es., si veda la Figura 2.3, il pannello in basso a destra).

La Figura 2.4 mostra la statistica Q , calcolata separatamente per i due indici di correlazione, come funzione degli stessi fattori descritti precedentemente. Questa volta le due variabili X e Y sono soggette a un faking in direzioni opposte (es., X è corrotta da un processo di faking good, mentre Y è perturbata da un processo faking bad, o vice versa). Da una piccola ispezione della Figura 2.4, è chiaro che le curve di questo grafico generale sono le stesse della Figura 2.3 ma speculari. In generale, per il modello opposto di faking con correlazioni originali non negative, l'indice Q di Cohen assume valori negativi. Per esempio, data una correlazione di Spearman non manipolata, $\rho_d = 0.6$ e un modello di faking average con una percentuale di fakers del 20%, osserviamo un bias dovuto al faking di $Q = -0.6$. Infine, le figure dalla 2.5 alla 2.8 offrono ulteriori rappresentazioni dettagliate con un ingrandimento in specifiche combinazioni di livelli di fattori per evidenziare alcune differenze rilevanti tra i due indici di correlazione. In sintesi, l'analisi grafica mostra chiaramente come un processo di risposta faking può modificare sostanziosamente il valore di correlazione della popolazione originale incrementandolo, o diminuendolo, in base alla quantità e qualità del meccanismo di perturbazione. Questo mostra chiaramente come le manipolazioni faking possano avere un impatto, in una maniera prevedibile, sulla correlazione della popolazione mista.

2.7 Studio simulativo

L'analisi a livello di popolazione rappresenta una maniera elegante ed efficiente per studiare l'impatto del faking su una relazione bivariata di variabili discrete con un intervallo limitato. In ogni caso, una delle limitazioni principali di questo tipo di analisi è che non prende in considerazione l'effetto dovuto alla dimensione del campione nel valutare il meccanismo di faking. Per superare questa mancanza, nella sezione seguente descriveremo uno studio simulativo il quale estende i risultati principali dell'analisi a livello di popolazione considerando fattori addizionali come la dimensione del campione e diverse numerosità di livelli di risposta per le variabili discrete coinvolte nella correlazione.

2.7.1 Disegno della simulazione

Abbiamo usato la versione campionaria q dell'indice di dimensione dell'effetto Q come variabile dipendente principale del disegno di simulazione:

$$q = \frac{1}{2} \log \frac{1 + r_m}{1 - r_m} - \frac{1}{2} \log \frac{1 + r_d}{1 - r_d} \quad (2.15)$$

dove r_d e r_m sono le correlazioni campionarie calcolate, rispettivamente, secondo le formule 2.1 e 2.3. Inoltre sei fattori sono stati fatti variare sistematicamente in un disegno sperimentale completo:

- a. la numerosità campionaria (N), in quattro livelli: 20, 50, 100, 1000;
- b. la proporzione di fakers nel campione perturbato (A), in dieci livelli: 0, .10, ..., .90;
- c. il tipo di modello di faking (M), in due livelli: faking *slight* e faking *extreme*;
- d. la direzione del faking (DIR), in due livelli: stessa (*same*) e opposta (*opposite*).
Una perturbazione di tipo *same* è ottenuta quando le due variabili nel pattern fake sono soggette alla stessa manipolazione faking good (oppure entrambe faking bad). Una perturbazione di tipo *opposite* richiede che nel pattern fake la prima variabile sia soggetta a faking good e la seconda a faking bad (oppure viceversa).
- e. il numero dei livelli di risposta per le variabili discrete (V), in 3 livelli: 2, 5, 7 opzioni di risposta;
- f. la correlazione nella popolazione non corrotta (R_d), in sei livelli: -.9, -.6, -.3, 0, .3, .6, .9.

Definiamo con $n, \alpha, m, dir, v,$ e ρ_d i distinti livelli dei fattori N, A, M, DIR, V e R_d , rispettivamente. I seguenti passaggi procedurali sono stati ripetuti 2,000 volte per ognuna delle 3,360 = $4 \times 10 \times 2 \times 2 \times 3 \times 7$ combinazioni dei livelli dei fattori $(n, \alpha, m, dir, v, \rho_d)$ del disegno di simulazione:

1. Generazione di un dataset bivariato di dati grezzi \mathbf{D} con dimensione n secondo la correlazione della popolazione non corrotta (ρ_d). La generazione dei dati è stata eseguita usando una procedura MC standard basata su dati normali multivariati (Kaiser & Dickman, 1962). Più specificatamente, ogni riga di \mathbf{D} è campionata dalla distribuzione normale bivariata $\phi(0, \Sigma_d)$, dove Σ_d è la matrice di correlazioni con gli elementi non sulla diagonale settati a ρ_d

2. Discretizzazione di \mathbf{D} sulla base di $v - 1$ differenti soglie usando il ben noto metodo descritto da Jöreskog e Sörbom (1996).¹
3. Per ogni osservazione $s = 1, \dots, n$ viene estratto un valore binario a sulla base della proporzione di fakers α . Se $a = 1$, allora s -esima riga di \mathbf{D} viene sostituita da una nuova riga, (riga di sostituzione, ottenuta sulla base del modello di faking m con direzione dir). Al contrario, se $a = 0$, viene mantenuta la riga s -th di \mathbf{D} . Questa procedura porta a una matrice mista \mathbf{X} .
4. Vengono calcolati e salvati gli effetti campionari q per i due indici di correlazione (Pearson e Spearman).

L'intera procedura è stata utilizzata per generare un totale di $6,720,000 = 2,000 \times 4 \times 10 \times 2 \times 2 \times 3 \times 7$ matrici corrotte, matrici non corrotte e anche coppie di indici di effect size². Alcuni dettagli tecnici relativi a principi di ottimizzazione basati sul calcolo parallelo che sono stati implementati nella simulazione MC sono riportati in Appendice A.

2.8 Risultati

Come atteso, i risultati MC convergono a quelli della popolazione quando si analizzano gli esiti ottenuti sui grandi campioni. Però, le prestazioni complessive dei due indici di associazione sono differenti quando si analizzano campioni piccoli. In particolare, la correlazione non parametrica ha prestazioni superiori rispetto alla correlazione prodotto-momento di Pearson (vedasi Figura 2.9) poiché le sue stime sono più vicine a quelle della popolazione anche nel caso di campioni molto piccoli (es., $n = 20$).

Le tabelle 2.1, 2.2, 2.3 riportano i valori della mediana della statistica q campionaria per i modelli di faking slight ed extreme, rispettivamente, con la stessa direzione. Come per l'analisi a livello di popolazione, la dimensione dell'effetto a livello campionario è più influenzata dalla presenza di faking extreme, e la magnitudine di questo effetto dipende chiaramente dal valore di correlazione originale.

Anche il fattore DIR modula il pattern osservato della statistica q . In particolare, la stessa direzione di faking per entrambe le variabili aumenta positivamente (in

¹Nello studio simulativo tutti i dataset continui \mathbf{D} sono stati discretizzati in variabili simmetricamente distribuite con i seguenti valori di massa: 0.5, 0.5 per $v = 2$; 0.06, 0.25, 0.38, 0.25, 0.06, per $v = 5$; 0.02, 0.09, 0.23, 0.31, 0.23, 0.09, 0.02, per $v = 7$.

²Nei passaggi 3 e 5, il campione generato, rispettivamente \mathbf{D} oppure \mathbf{X} , è stato ricampionato nel caso in cui entrambi gli indici di correlazione non avessero dato un valore valido (p.e., il fatto che una colonna dei dati avesse zero varianza non permette il calcolo della correlazione). Nonostante questo, una piccola percentuale dei dati è stata comunque scartata, circa lo 0.035% dei 6,720,000 campioni totali.

Tabella 2.1: Dimensione dell'effetto q come funzione della percentuale di fakers, correlazione originale e tipo di modello di faking. Caso di variabili binarie, $V = 2$, con la stessa direzione di faking.

| R_d | Slight | | | | | Extreme | | | | | | |
|-------|--------|-------|-------|-------|-------|---------|-------|-------|-------|-------|-------|-------|
| | 10% | 20% | 40% | 60% | 80% | 90% | 10% | 20% | 40% | 60% | 80% | 90% |
| -0.9 | 0.269 | 0.440 | 0.672 | 0.831 | 0.955 | 0.986 | 0.263 | 0.436 | 0.676 | 0.832 | 0.946 | 0.992 |
| -0.6 | 0.146 | 0.260 | 0.430 | 0.558 | 0.659 | 0.709 | 0.145 | 0.255 | 0.430 | 0.556 | 0.665 | 0.700 |
| -0.3 | 0.110 | 0.201 | 0.347 | 0.457 | 0.557 | 0.594 | 0.110 | 0.203 | 0.349 | 0.460 | 0.555 | 0.594 |
| 0 | 0.089 | 0.167 | 0.298 | 0.400 | 0.483 | 0.523 | 0.090 | 0.167 | 0.298 | 0.396 | 0.474 | 0.521 |
| 0.3 | 0.075 | 0.141 | 0.255 | 0.359 | 0.432 | 0.491 | 0.075 | 0.145 | 0.256 | 0.352 | 0.428 | 0.481 |
| 0.6 | 0.064 | 0.121 | 0.233 | 0.316 | 0.398 | 0.451 | 0.062 | 0.123 | 0.224 | 0.317 | 0.394 | 0.432 |
| 0.9 | 0.047 | 0.098 | 0.192 | 0.278 | 0.343 | 0.372 | 0.049 | 0.102 | 0.189 | 0.263 | 0.363 | 0.350 |

Tabella 2.2: Dimensione dell'effetto q come funzione della percentuale di fakers, correlazione originale e tipo di modello di faking. Caso di variabili binarie, $V = 5$, con la stessa direzione di faking.

| R_d | Slight | | | | | | | | | Extreme | | | | |
|-------|--------|-------|--------|--------|--------|--------|-------|-------|-------|---------|--------|--------|--|--|
| | 10% | 20% | 40% | 60% | 80% | 90% | 10% | 20% | 40% | 60% | 80% | 90% | | |
| -0.9 | 0.433 | 0.654 | 0.873 | 0.931 | 0.839 | 0.712 | 0.753 | 1.078 | 1.436 | 1.630 | 1.627 | 1.497 | | |
| -0.6 | 0.213 | 0.353 | 0.510 | 0.546 | 0.469 | 0.363 | 0.427 | 0.681 | 0.991 | 1.153 | 1.122 | 0.987 | | |
| -0.3 | 0.138 | 0.241 | 0.359 | 0.382 | 0.307 | 0.218 | 0.306 | 0.507 | 0.780 | 0.911 | 0.852 | 0.699 | | |
| 0 | 0.101 | 0.178 | 0.268 | 0.277 | 0.201 | 0.118 | 0.235 | 0.404 | 0.641 | 0.738 | 0.658 | 0.480 | | |
| 0.3 | 0.078 | 0.137 | 0.198 | 0.189 | 0.110 | 0.029 | 0.190 | 0.340 | 0.534 | 0.584 | 0.456 | 0.261 | | |
| 0.6 | 0.057 | 0.097 | 0.132 | 0.101 | 0.001 | -0.076 | 0.153 | 0.274 | 0.418 | 0.411 | 0.223 | -0.008 | | |
| 0.9 | 0.023 | 0.026 | -0.008 | -0.096 | -0.238 | -0.324 | 0.108 | 0.179 | 0.209 | 0.083 | -0.230 | -0.499 | | |

Tabella 2.3: Dimensione dell'effetto q come funzione della percentuale di fakers, correlazione originale e tipo di modello di faking. Caso di variabili binarie, $V = 5$, con la stessa direzione di faking.

| R_d | Slight | | | | | Extreme | | | | | | |
|-------|--------|--------|--------|--------|--------|---------|-------|-------|-------|--------|--------|--------|
| | 10% | 20% | 40% | 60% | 80% | 90% | 10% | 20% | 40% | 60% | 80% | 90% |
| -0.9 | 0.475 | 0.701 | 0.926 | 0.980 | 0.893 | 0.780 | 0.900 | 1.246 | 1.621 | 1.782 | 1.727 | 1.562 |
| -0.6 | 0.214 | 0.358 | 0.507 | 0.544 | 0.468 | 0.373 | 0.499 | 0.777 | 1.105 | 1.249 | 1.159 | 0.989 |
| -0.3 | 0.137 | 0.236 | 0.348 | 0.372 | 0.297 | 0.211 | 0.357 | 0.590 | 0.882 | 0.991 | 0.873 | 0.691 |
| 0 | 0.097 | 0.172 | 0.255 | 0.260 | 0.181 | 0.107 | 0.278 | 0.470 | 0.724 | 0.800 | 0.652 | 0.449 |
| 0.3 | 0.071 | 0.128 | 0.177 | 0.165 | 0.083 | 0.010 | 0.221 | 0.395 | 0.595 | 0.622 | 0.432 | 0.210 |
| 0.6 | 0.049 | 0.078 | 0.103 | 0.058 | -0.043 | -0.120 | 0.180 | 0.317 | 0.462 | 0.411 | 0.164 | -0.079 |
| 0.9 | 0.002 | -0.021 | -0.094 | -0.213 | -0.356 | -0.454 | 0.123 | 0.197 | 0.183 | -0.005 | -0.358 | -0.633 |

media) la correlazione finale, mentre la direzione *opposta* di faking riduce (in media) la correlazione finale perturbata. Questi effetti sono coerenti in tutte le combinazioni di fattori coinvolti nel disegno di simulazione, a meno che non si considerino variabili politomiche con correlazioni originali estremamente positive (es., 0.9) nelle quali si può osservare un effetto di q mediamente negativo (tabelle 2.2, 2.3, ultima riga).

Per il fattore V , osserviamo un effetto di interazione con la proporzione di fakers A . In particolare, per le variabili dicotomiche osserviamo una relazione monotonica positiva tra la quantità di fakers nel campione e l'indice q . Al contrario, una relazione concava emerge quando si considerano variabili politomiche ($V = 5$ or $V = 7$). Infine, per le variabili politomiche con correlazioni originali negative, il più grande effetto del faking è osservato quando ci sono circa il 60% di fakers nel campione ($q > 0.4$).

Un altro effetto interessante può essere osservato per il fattore R_d . Nel caso di manipolazioni faking, le correlazioni originariamente basse (es., $|\rho_d| \leq 0.3$) possono essere affette da bias sia in termini di magnitudine osservata sia come segno della correlazione. Per esempio, una correlazione originale $\rho_d = -0.3$ può essere trasformata in una correlazione finale positiva quando un effetto medio-grande di q è presente (p.e., tabella 2, le celle corrispondenti al modello estremo con una percentuale di fakers del 20%).

Infine, alcuni importanti risultati riguardano l'effetto della dimensione del campione N oppure la proporzione di fakers sulla variabilità dell'indice q . Come atteso, una relazione negativa si verifica tra N e la varianza di q : più piccola è la dimensione del campione, più grande è la variabilità della statistica q (vedasi Figura 2.10). Al contrario, si osserva una associazione positiva tra A e q : più grande è la proporzione di fakers nel campione, più ampia è la varianza di q (vedasi Figura 2.10).

Data la grande regolarità della struttura del disegno dello studio di simulazione, risultati completamente simmetrici sono stati osservati per la direzione di faking opposta (faking bad).

2.9 Breve discussione e conclusioni

La motivazione principale di questo lavoro era di mostrare l'impatto della minaccia dei dati fake quando si analizzano misure self-report. Abbiamo limitato lo studio alla valutazione di dati fake a due semplici misure bivariate di associazione: la correlazione prodotto-momento e la correlazione a ranghi di Spearman. Abbiamo mostrato che il paradigma SGR può essere usato per predire il comportamento degli indici di correlazione sotto le perturbazioni di tipo fake. In particolare, abbiamo presentato un'estensione della rappresentazione SGR la quale si basa su un modello misto, il do-

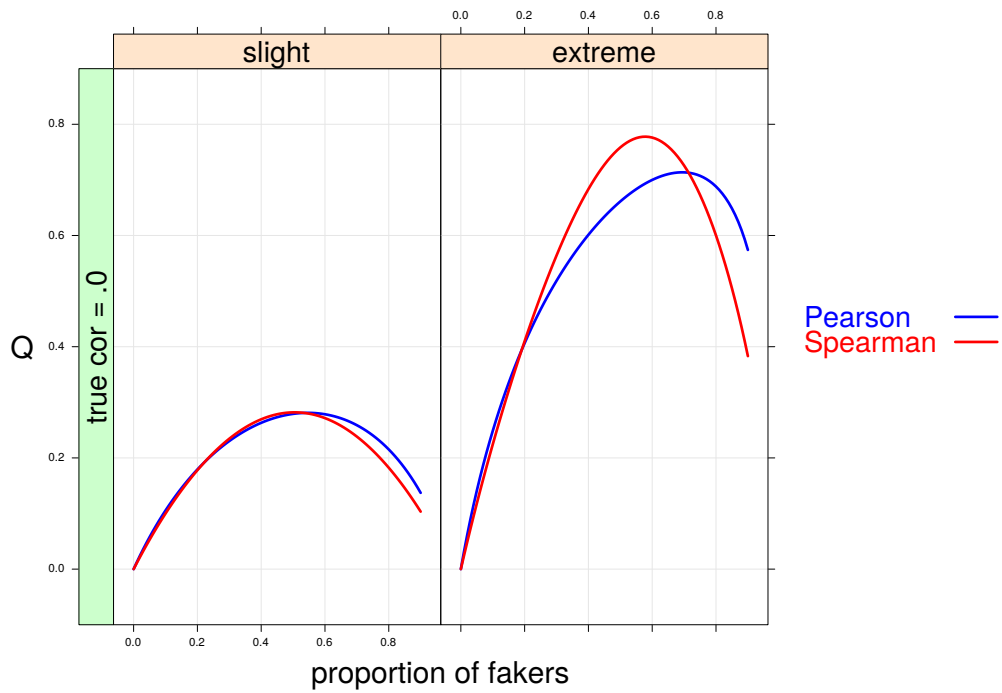


Figura 2.5: La statistica Q di Cohen come funzione della proporzione di fakers e modello di faking. In questo esempio, sono mostrati $\rho_d = 0$, $v = 5$ e stessa direzione di faking.

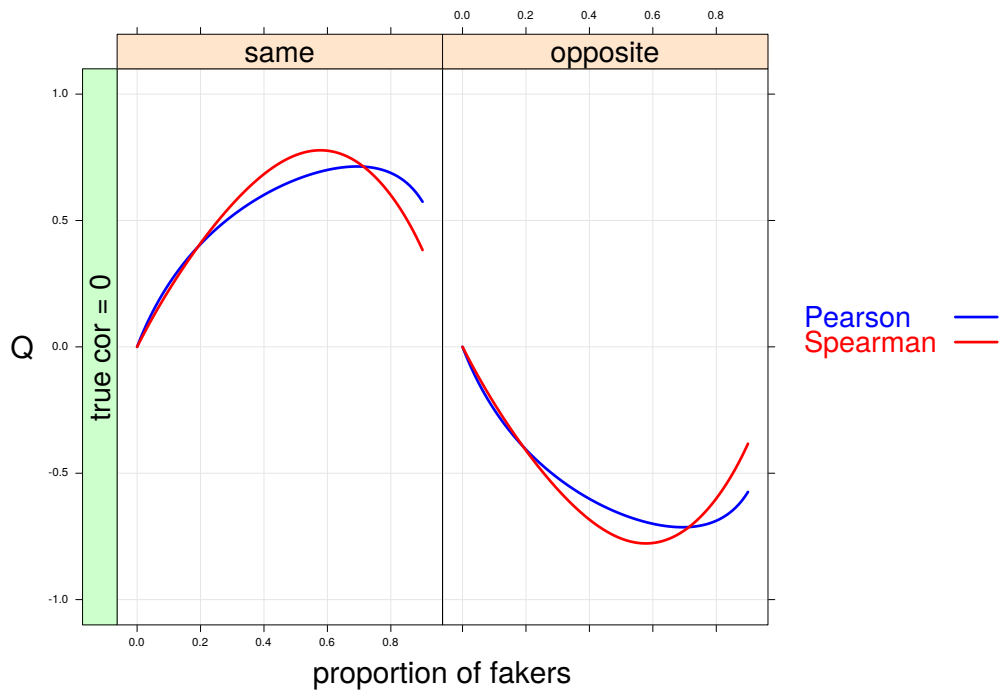


Figura 2.6: La statistica Q di Cohen come funzione della proporzione di fakers e direzione di faking nelle due variabili. In questo esempio, sono mostrati $\rho_d = 0$, $v = 5$ e un modello di faking extreme.

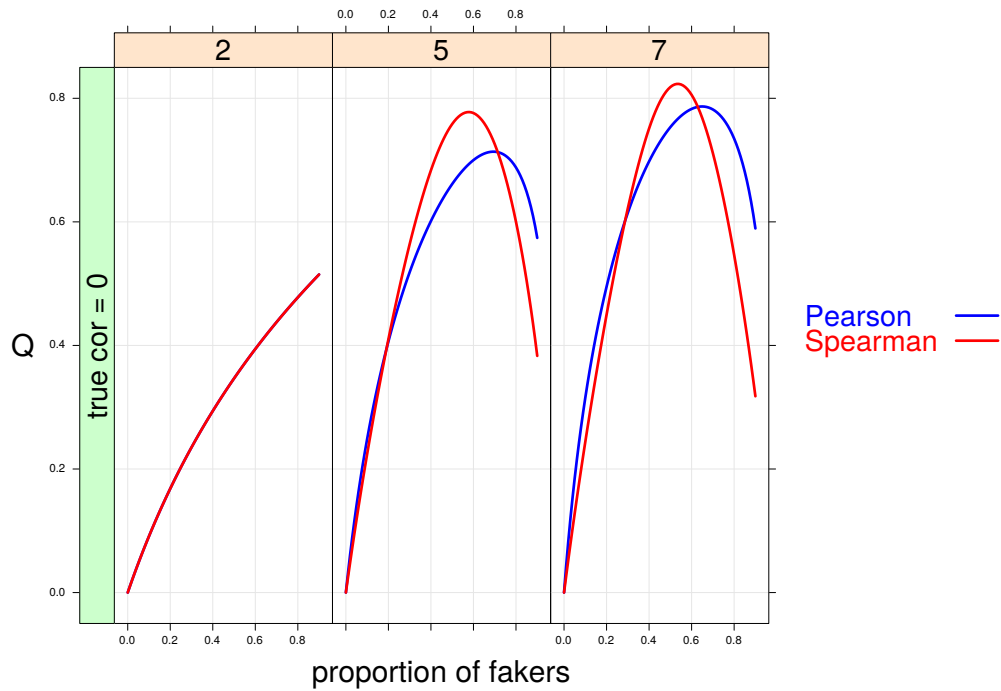


Figura 2.7: La statistica Q di Cohen come funzione della proporzione di fakers e 3 numerosità dei livelli delle variabili discrete ($v = 2, 5, 7$). In questo esempio, sono mostrati $\rho_d = 0$, un modello di faking extreme con la stessa direzione. Nota: nel primo riquadro ($v = 2$) i due indici di correlazione sono completamente sovrapposti.

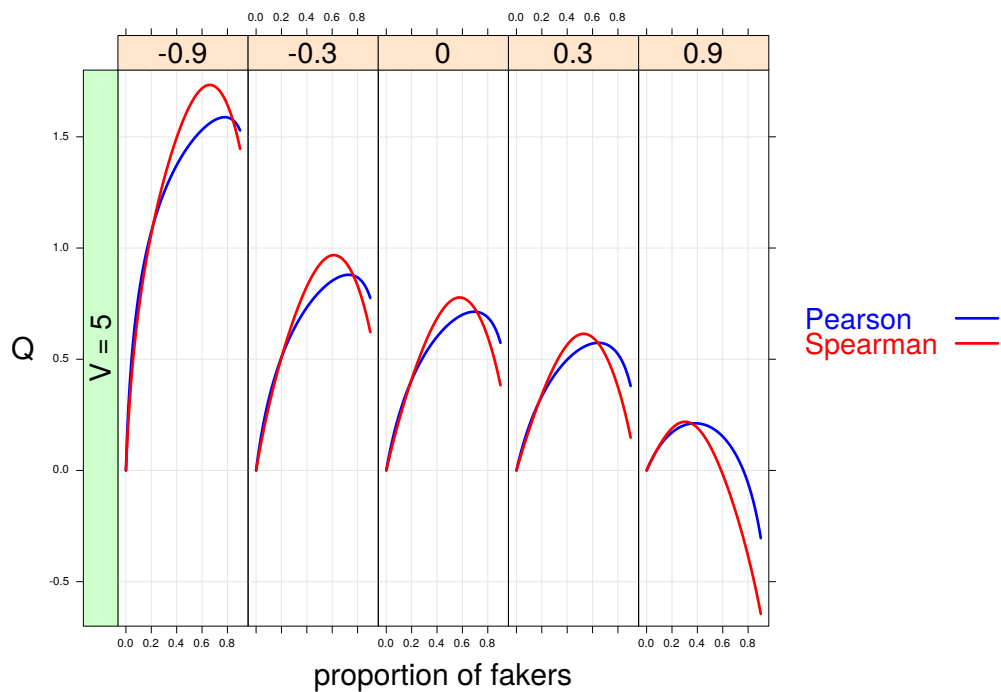


Figura 2.8: La statistica Q di Cohen come funzione della proporzione di fakers e la correlazione non corretta a 5 livelli ($\rho_d = -0.9, -0.3, 0, 0.3, 0.9$). In questo esempio, sono mostrati $v = 5$, modello di faking extreme con la stessa direzione.

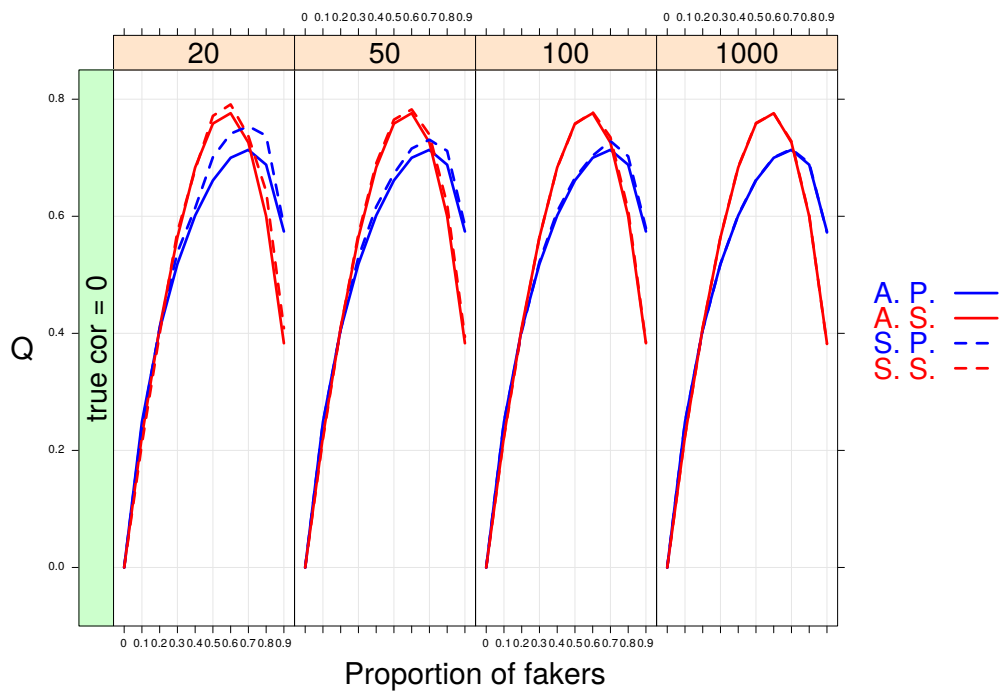


Figura 2.9: La statistica Q di Cohen come funzione della proporzione di fakers e dimensione campionaria in 4 livelli ($n = 25, 50, 100, 1000$). In questo esempio, sono mostrati $\rho_d = 0$, $v = 5$, e un modello di faking extreme con la stessa direzione. A.P. e A.S., rispettivamente, stanno per correlazione di Pearson oppure Spearman calcolate a livello di popolazione. S.P. e S.S., stanno, rispettivamente, per correlazione di Pearson e correlazione di Spearman stimate tramite procedura Monte Carlo (MC).

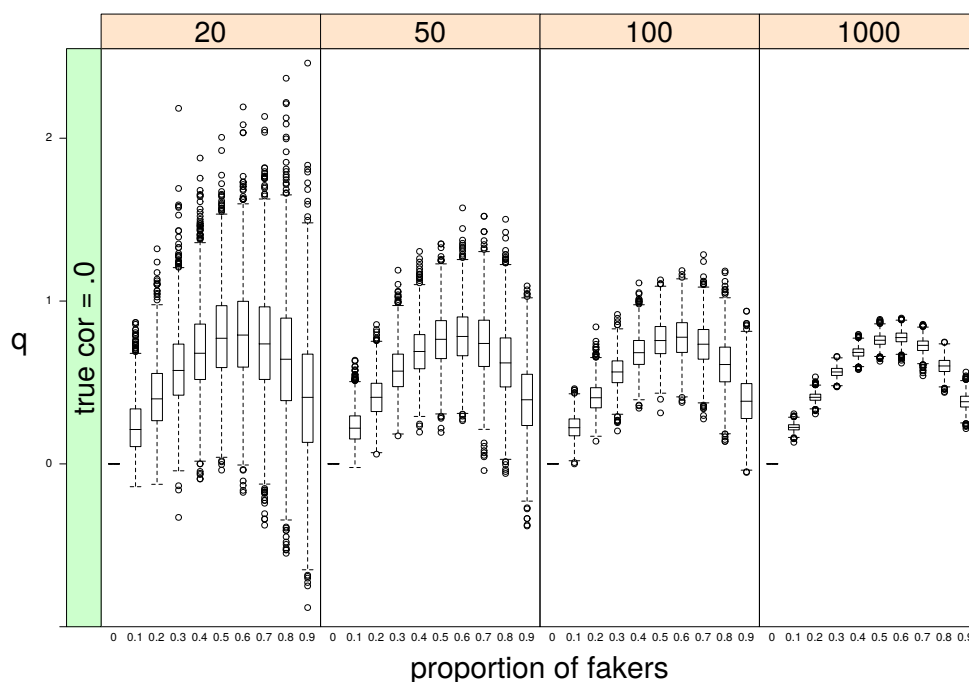


Figura 2.10: Boxplot della statistica campionaria q come funzione della proporzione di fakers nel campione e della dimensione del campione a 4 livelli ($n = 20, 50, 100, 1000$). In questo esempio, è stata calcolato solo q di Spearman; sono mostrati $\rho_d = 0$, $v = 5$, e un modello di faking extreme con la stessa direzione.

rebbe essere più coerente con la modalità degli individui di manipolare le risposte self-report dato che può tenere conto di differenze individuali nello stile di menzogna. Abbiamo inoltre fornito una rappresentazione del problema del faking a livello di popolazione modellizzando direttamente le distribuzioni di probabilità congiunta a livello di popolazione.

Il risultato principale della nostra analisi SGR è che è possibile, sotto certe condizioni di faking, sia aumentare che diminuire una correlazione originale non corrotta, in una maniera prevedibile e sistematica. Da una prospettiva applicativa, questi risultati generali hanno molte rilevanti implicazioni. In particolare, vorremo comunicare come una correlazione misurata potrebbe essere dovuta a una falsa correlazione positiva (correlazione spuria) risultante da manipolazioni di tipo faking strutturante e non a una reale connessione tra le variabili. Oltretutto, se l'effetto del faking è sufficientemente grande, la correlazione potrebbe anche sembrare significativa a livello campionario. Vogliamo dire che, stabilita una qualsiasi dimensione dell'effetto dovuta al faking Q_f e l'associata correlazione r_f , è facile calcolare la dimensione campionaria minima n_c per cui tale correlazione sia significativa. Al contrario, correlazioni non significative a livello campionario, potrebbero non mostrare le reali associazioni poiché corrotte da

qualche forma di faking destrutturante. I risultati dello studio a livello di popolazione e della simulazione MC hanno sottolineato che l'impatto del faking (sull'associazione tra due variabili) è principalmente influenzato dallo specifico modello di faking adottato per manipolare i dati non corrotti. In altre parole, l'impatto dipende principalmente da quanto il processo di faking va a cambiare le medie marginali delle due variabili. Come atteso, il modello di faking extreme ha un impatto più forte sulla media se comparato al modelli slight, ciò comporta anche un effetto più grande sulla varianza delle due variabili e di conseguenza sulla correlazione finale osservata.

I più grandi pericoli derivanti dagli effetti delle manipolazioni di tipo faking sono per le piccole o medie correlazioni, dove anche piccoli effetti del faking possono drasticamente cambiare il p -value osservato nella correlazione tra le due variabili. Inoltre, come le correlazioni estreme sono sensibili alle manipolazioni perché potrebbero diventare correlazioni medie, le correlazioni piccole potrebbero essere affette dalle manipolazioni faking in entrambe le direzioni (mostrando sia relazioni positive che negative) nella correlazione empirica. Quest'ultima osservazione implica che le manipolazioni faking potrebbero facilmente invertire il segno di una correlazione non corrotta (*effetto segno del faking*).

Come in altri studi Monte Carlo, la nostra ricerca assume alcuni assunti che potrebbero portare a una validità esterna più bassa come, per esempio, l'assunzione che i valori delle soglie nel modello generativo siano considerate costanti nelle variabili e simmetriche attorno alla media della distribuzione bivariata standardizzata. Sfortunatamente, questa restrizione chiaramente limita il campo di utilizzo empirico del processo di faking che può essere imitato utilizzando il nostro approccio. In ogni caso, non è difficile modificare il processo di campionamento per garantire l'utilizzo di distribuzioni asimmetriche per il processo di generazione dei dati secondo il modello di generazione originale. A questo riguardo, mettiamo a disposizione il codice completo in linguaggio R per poter eseguire analisi a livello di popolazione che tengano conto di diverse proprietà distribuzionali della distribuzione congiunta del modello originale non perturbato (p.e., distribuzioni asimmetriche e con elevata curtosi).

Inoltre, una naturale estensione della modellizzazione SGR dovrebbe riguardare le differenze individuali nella modellizzazione della distribuzione di sostituzione e andare oltre la semplice rappresentazione con due categorie (onesti vs manipolatori) fornita in questo capitolo. Più specificatamente, potremmo imitare diversi parametri del faking sulla base di alcune caratteristiche dei soggetti o caratteristiche nel modello misto. In questa generalizzazione potremmo utilizzare c differenti categorie di pattern faking

come segue:

$$s_{xy} = (1 - \alpha^*)p_{xy} + \sum_{l=1}^c \alpha_l z_{xy}^{(l)}, \quad (x, y) \in \{1, 2, \dots, v\}^2 \quad (2.16)$$

dove s_{xy} è la probabilità del modello misto di osservare le variabili $X = x$ e $Y = y$. Nell'equazione 2.16, il parametro $\alpha_l \in [0, 1]$ definisce la proporzione di fakers di tipo l nella popolazione, tale che

$$0 \leq (\alpha^* = \sum_{l=1}^c \alpha_l) \leq 1.$$

Per esempio, potremmo modellare tre differenti tipi ($c = 3$) di comportamenti di faking (es., slight, average, ed extreme) nella popolazione, con la corrispettiva proporzione. In questa maniera, potremmo utilizzare informazioni aggiuntive riguardo le caratteristiche dei rispondenti (es., misure di desiderabilità) per settare i parametri di faking del modello di sostituzione come funzioni delle stesse.

Infine, per applicazioni empiriche ipotesi specifiche riguardo la modellizzazione (sia il modello vero che la distribuzione di sostituzione) potrebbero essere derivate da studi precedenti, seguendo lo stesso ragionamento che si potrebbe trovare nella prospettiva Bayesiana, per esempio, il settaggio di distribuzioni prior sulla base di precedenti osservazioni empiriche. Similmente, anche assunzioni specifiche riguardo la natura dell'associazione potrebbero essere estratte da norme pubblicate in specifici strumenti di misura (es., quando uno specifico test psicologico viene pubblicato o aggiornato), o da aspettative esplicite derivanti da teorie della personalità o della motivazione (es., impostare entrambi i parametri della distribuzione di sostituzione dopo averli appropriatamente formalizzati e identificati).

In ogni caso, è importante evidenziare e sottolineare come il faking costituisca solo uno di molti esempi di manipolazione di dati che potrebbero avere effetti sui risultati statistici nelle misure self-report (es., copiare le risposte o provare ad indovinarle, rispondere casualmente, invertire le risposte o preferire quelle estreme). Perciò studi futuri dovrebbero inoltre considerare l'interazione tra queste differenti fonti di manipolazione dei dati sui risultati delle analisi statistiche.

Capitolo 3

Identificare pattern di risposte manipolate

3.1 Introduzione

Quando si utilizzano questionari in situazioni particolari (es. selezione del personale, dispute legali) oppure in contesti sensibili (es. comportamenti a rischio oppure illeciti), sia che i questionari siano di abilità o conoscenza (in cui esiste una risposta "corretta") oppure di opinioni, personalità o autodichiarazioni in generale (dove non ha senso parlare di risposta "corretta", ma al massimo di sincera), i rispondenti hanno talvolta la tendenza a imbrogliare, copiare o manipolare le risposte. Per porre rimedio a ciò gli psicometristi hanno da sempre cercato di individuare procedure per limitare la manipolazione dei dati oppure di identificare i rispondenti non sinceri al fine di escluderli (Ziegler, MacCann, & Roberts, 2012).

L'obiettivo di questo capitolo è mostrare come non sia sempre facile identificare i soggetti manipolatori sulla base del loro pattern di risposte, nel caso in cui stiano rispondendo ad un questionario su scala ordinale. Allo stesso modo non è scontata l'identificazione dei fattori che vadano ad aumentare, o mitigare, gli effetti sui risultati statistici della manipolazione faking sui dati osservati.

Il capitolo è organizzato come segue, nella prima sezione è descritta una rassegna relativa alle manipolazioni dei dati indagate nella letteratura attraverso simulazioni Monte Carlo. Lo scopo di questa prima parte è quello di mettere in evidenza come la metodologia SGR sia in grado di riprodurre la maggior parte degli schemi e dei modelli di manipolazione descritti nel corso degli anni in letteratura. Nella seconda parte è descritta una simulazione Monte Carlo, dove viene analizzata l'efficacia di alcuni indici Person Fit Statistics (PFS) nell'identificare eventuali soggetti fakers nel caso in cui le risposte siano di tipo politomico-ordinale. Infine, nell'ultima breve sezione di questo

capitolo viene presentato un "caso critico", analizzato attraverso un approccio statistico basato sul concetto di likelihood, che mette in evidenza come in generale i PFS non riescano efficacemente a distinguere i fakers dai risponditori onesti.

3.2 Caratteristiche delle perturbazioni dei pattern in simulazioni Monte Carlo

Un insieme di risposte ad un test è detto aberrante quando il soggetto, ad esempio, non è stato sincero, ha intenzionalmente mentito, copiato delle risposte, risposto casualmente, risposto distrattamente. Quando succede ciò possiamo osservare due distinti scenari: a) l'insieme di risposte date è coerente con il livello di costrutto stimato da un modello b) le risposte non sono coerenti tra loro (Levine & Rubin, 1979), o emergono pattern anomali / inusuali / infrequenti la cui validità può essere discutibile e necessita approfondimento (Meijer & Sijtsma, 2001).

Immaginiamo un test di abilità matematica in cui il soggetto sbaglia delle semplici addizioni, ma al tempo stesso riesce a risolvere degli integrali molto difficili. Probabilmente il suo punteggio totale finale risulterà sopra la media, sebbene il pattern delle risposte potrebbe risultare in parte incoerente. Viceversa, uno studente poco brillante che copia alcune risposte di media difficoltà dal suo vicino molto bravo, probabilmente otterrà un punteggio totale nella media, ma questo non rappresenterà il suo reale livello. Un altro esempio è dato dalla situazione in cui una persona risponde casualmente alle domande di un inventario di personalità, producendo così pattern di risposta incoerenti. Come ultimo esempio possiamo pensare a una persona che vuole essere assunta ad ogni costo e, rispondendo a un questionario, cerca di mostrarsi più estroversa e coscienziosa di quanto in realtà lo sia, convinta che queste caratteristiche siano desiderabili dal datore di lavoro.

Una metodologia di aberrazione è una tecnica, solitamente di tipo deterministico o probabilistico, che è utilizzata per produrre l'alterazione di uno a più pattern di risposte a un test. Alcuni autori hanno definito dei modelli matematici i quali, manipolando un pattern di risposte di un test oppure i parametri del modello generativo, tentano di simulare il comportamento di una persona che sta alterando, più o meno intenzionalmente, le risposte. Tutte le aberrazioni citate nella prossima sezione cercano di imitare un particolare comportamento dei soggetti. La seguente review è limitata principalmente ai casi dicotomici e nello specifico sono riportate per esteso le definizioni delle metodologie di aberrazione di pattern di risposta utilizzate in simulazioni Monte Carlo, così come descritte dai rispettivi autori. Le definizioni sono poi riassunte in tabella 3.1. La distinzione tra dicotomico e politomico dipende dal modo in cui è stata

fatta la codifica della risposta, ad esempio, in un item a risposta multipla, in cui solo una risposta è esatta ed essa viene codificata come 1, mentre le altre risposte vengono codificate come 0, è stato considerato come item dicotomico. Generalmente gli item codificati come dicotomici appartengono a questionari di abilità a risposta multipla, questo spiega perché la probabilità che la risposta casuale sia corretta è solitamente fissata a $\frac{1}{v}$, dove v è il numero di possibili risposte all'item. Al contrario, gli item politomici, sono usualmente rappresentati da scale Likert, utilizzate nei questionari per misurare, ad esempio, gli atteggiamenti o la personalità.

3.2.1 Codifica dicotomica della risposta

Qui di seguito sono riportate 26 definizioni di manipolazioni utilizzate da diversi autori per simulare il comportamento aberrante dei soggetti su item codificati con risposte dicotomiche. Queste definizioni sono state in seguito aggregate in base alla loro logica di aberrazione per cercare di ridurre il numero. Nella tabella 3.1 sono riassunte le aberrazioni studiate. Attenzione, in letteratura talvolta vengono utilizzati nomi diversi corrispondenti allo stesso tipo di manipolazione e, specularmente, manipolazioni strutturalmente differenti possono essere indicate con lo stesso nome.

Levine e Rubin (1979, pg. 277) illustrano due tipologie di aberrazione: *spuriously high* e *spuriously low*. *Spuriously high* è stato ideato per simulare un comportamento di un soggetto che ottiene un punteggio fittiziamente più alto di quello che avrebbe ottenuto senza manipolare le risposte perché, ad esempio, ne ha copiate il 20%. Il primo passaggio consiste nel generare il pattern delle risposte di un soggetto onesto. Come secondo passaggio il 20% delle risposte veniva campionato senza reinserimento. Come ultimo passaggio, le risposte precedentemente campionate venivano segnate come corrette, indipendentemente dal fatto che prima fossero corrette oppure no. La seconda tipologia è chiamata *spuriously low*. Questa seconda metodologia di aberrazioni è stata pensata per simulare un comportamento di un soggetto che ottiene un punteggio test fittiziamente più basso perché il soggetto ha risposto casualmente ad alcune risposte, magari per stanchezza o inattenzione. I primi due passaggi si ripetono come in *spuriously high*. Nel terzo passaggio la risposta viene modificata su base probabilistica e il soggetto ha una probabilità di $\frac{1}{v}$, di ottenere una risposta corretta e $\frac{v-1}{v}$ di ottenere quella sbagliata. Se con la *spuriously high* il punteggio dopo la manipolazione è sicuramente maggiore o uguale al punteggio onesto, nella *spuriously low* il punteggio test ottenuto dalla manipolazione può risultare sia superiore che inferiore a quello senza manipolazione (la direzione del risultato dipende dal punteggio onesto e dagli esiti del campionamento e dalla manipolazione probabilistica).

Drasgow (1982, pg. 301) utilizza il termine *spuriously high* per indicare una manipolazione che, selezionato un soggetto di bassa abilità e scegliendo in modo casuale un certo numero di item, segna la risposta corretta, indipendentemente dal fatto che lo fosse o meno. Invece, usa il termine *spuriously low* per indicare una manipolazione che, selezionato un soggetto di abilità alta e scegliendo in modo casuale un certo numero di item, segna la risposta corretta con una probabilità di $\frac{1}{v}$ o scorretta con una probabilità di $\frac{v-1}{v}$. Quindi Drasgow (1982) seleziona gli item con lo stesso modo di Levine e Rubin (1979), ma per i soggetti da manipolare utilizza un criterio di selezione. Questo criterio è scegliere soggetti con un punteggio test basso e, attraverso una procedura deterministica, ottenere un punteggio pari o superiore a quello di partenza; o in alternativa scegliere dei soggetti con un punteggio alto e con una procedura probabilistica diminuire il punteggio.

Rudner (1983, pg. 208) utilizza ancora gli stessi nomi ma cambia le definizioni. Con *spuriously high* intende il cambiamento deterministico di un certo numero di risposte del soggetto da incorrette a corrette, mentre per *spuriously low* il cambiamento di un certo numero di risposte da corrette a incorrette. Questo autore non seleziona quindi le risposte casualmente, ma le campiona esclusivamente tra quelle incorrette (*spuriously high*) o corrette (*spuriously low*).

Drasgow, Levine, e Williams (1985, pg. 80) introducono il concetto di *spuriously high* per indicare una manipolazione che, fissata una percentuale di risposte da manipolare k e j numero di item del test, segna come corrette kj risposte estratte casualmente dal test, senza reinserimento, indipendentemente dall'abilità del soggetto. *Spuriously low*, invece è stato definito in modo leggermente più complesso e riflette un contesto dove l'individuo ha la propensione ad omettere le risposte. Dopo aver stimato q come percentuale di risposte omesse da un determinato soggetto, egli va a campionare una percentuale di k risposte, alla stessa maniera di *spuriously high*, e le segna come omesse con una probabilità pari a q oppure casuali con una probabilità di $\frac{1-q}{v}$.

Meijer, Molenaar, e Sijtsma (1994, pg. 114-115) definiscono il concetto di *cheaters*. Come primo step, si estrae un insieme di livelli di abilità dei soggetti (θ è il parametro che definisce l'abilità del soggetto secondo la rappresentazione dell'Item Response Theory) da una distribuzione normale. Nel secondo step vengono selezionati solo quei soggetti che risultano sotto la media e questi sono successivamente trasformati in *cheaters*. Questi autori usano anche la manipolazione *guessing*, nella quale si assume che il soggetto abbia risposto casualmente ad ogni risposta del test, con una probabilità di risposta corretta di $\frac{1}{v}$.

Meijer (1994, pg. 312) modifica leggermente la procedura per simulare il *cheating*, agendo non solo sui soggetti con abilità sotto la media ($\theta < 0$), ma estraendo i valori

delle abilità θ da una distribuzione normale.

Mentre i precedenti modelli di aberrazione erano utilizzati su test di conoscenza-abilità, Burnkrant e Harvey (2000, pg. 3) simulano uno scenario di test di personalità e introducono il concetto di *distortion*. Distortion è una manipolazione dove il soggetto, pur conoscendo la risposta che più lo rappresenta, vuole confondere il ricercatore e da la risposta opposta. Per simulare questo tipo di comportamento Burnkrant e Harvey (2000), dopo aver generato i pattern secondo modello onesto, invertono una certa percentuale di risposte scelte casualmente: 0 diventa 1 e viceversa. Inoltre usano il nome *totally-random* per indicare la condizione in cui il soggetto risponde completamente a caso. Estraendo da una distribuzione uniforme di numeri casuali assegnano la risposta 1 se il numero è maggiore di 0 e 0 se il numero estratto è minore di 0, praticamente è come se avessero campionato ogni risposta da una distribuzione Bernoulliana con $p = .5$.

Emons, Glas, Meijer, e Sijtsma (2003, pg. 469) interpretano il *cheating* come quel fenomeno associato a quei casi dove una persona per aumentare il proprio punteggio test, o impara a memoria le risposte più difficili oppure si siede vicino ad una persona con più alte capacità e, da questa, copia le risposte. Il pattern di risposte del soggetto cheater viene quindi da modello, ad eccezione delle risposte agli item con più alta difficoltà, le quali vengono settate come corrette (la percentuale di risposte manipolate è una variabile del disegno sperimentale). L'altra manipolazione che utilizzano è chiamata *inattention* e identifica quella situazione nella quale una persona per distrazione, oppure ansia, va a sottostimare la difficoltà di domande semplici e segna la risposta errata. Generando da modello il pattern del soggetto inattento, la probabilità di successo per gli item più facili viene settata a $\frac{1}{v}$.

Karabatsos (2003, pg. 283) utilizza ben cinque aberrazioni differenti. Cominciando dal *cheating*, dove genera un soggetto con un'abilità bassa ($-2 < \theta < -.5$) e gli assegna come corrette le risposte al 18% degli item più difficili, gli item con difficoltà δ maggiore di 1.5. Definisce anche il concetto di *creative*, soggetti che cercando una risposta originale o bizzarra, sbagliano le risposte agli item più facili: dopo aver generato i soggetti da modello segna come incorrette le risposte agli item appartenenti al 18% più facile del test, quelli con $\delta < -1.5$. La manipolazione *lucky guessing* serve a simulare un soggetto con una abilità inferiore alla media ($-2 < \theta < -.5$) il quale, non conoscendo le risposte più difficili, risponde a caso al 41% degli item più difficili ($\delta > .5$). Per tutti questi item viene estratta una probabilità di risposta corretta pari a $\frac{1}{v}$. Ci si aspetta che in questo modo il soggetto aumenti il punteggio atteso del test perché la probabilità di risposta corretta, assegnatagli agli item più difficili, è più alta di quella che avrebbe da modello. Karabatos utilizza anche la manipolazione chiamata *careless*, che descrive

il comportamento di quei soggetti con un'abilità sopra la media, $.5 < \theta < 2$, i quali, per distrazione, sbagliano a rispondere al 41% delle risposte più facili del test ($\delta < -.5$). A questi item più facili viene assegnata una possibilità di risposta corretta pari a $\frac{1}{2}$. Infine la manipolazione *randomly*, dove, per generare un soggetto le cui risposte sono completamente a caso, per ogni risposta del pattern aberrato è stato assegnata una probabilità di risposta corretta pari a $\frac{1}{v}$.

Una sola manipolazione è utilizzata da St-Onge, Valois, Abdous, e Germain (2009, pg. 316), *spuriously high*. Riprendono *spuriously high* descritto da Drasgow (1982), dopo aver generato i pattern secondo un modello e scelto i pattern da aberrare, spiegano i tre passaggi per aberrare il pattern: primo, gli item con una risposta errata vengono identificati; secondo, il 20% di queste risposte errate vengono estratte casualmente e identificate come quelle da modificare; terzo viene segnata una risposta corretta agli item selezionati.

St-Onge, Valois, Abdous, e Germain (2011, pg. 420-421) utilizzano le seguenti due aberrazioni sempre ispirandosi a Drasgow (1982). La prima manipolazione utilizzata è *cheating*, dove soggetti con un'abilità sotto la media ($\theta < 0$) vengono scelti per essere aberrati e le loro risposte errate identificate. A questo punto una percentuale di risposte viene segnata come corretta. La seconda manipolazione che utilizzano è *unlucky guessing*. In questa seconda manipolazione vengono selezionati soggetti con un'abilità sopra la media $\theta > 0$ e identificate le loro risposte corrette. A questo punto una certa percentuale di risposte corrette vengono ricodificate come errate con una probabilità di $\frac{v-1}{v}$.

Tra le 16 tipologie di aberrazione derivate da quelle presenti in letteratura, ne abbiamo selezionate 5 (la 2,3,5,6,7 della tabella 3.2), le quali sono le più affini alla metodologia SGR in quanto l'aberrazione risulta condizionata dalla risposta onesta. Dato questo sottogruppo di aberrazioni, sono state ricavate altre due tipologie di aberrazione non usate precedentemente in letteratura, la numero 1 e la 4 (tabella 3.1), che vanno strutturalmente ad aumentare il panorama delle manipolazioni fin ora prese in considerazione nella letteratura psicometrica.

Tabella 3.1: Manipolazioni per dati dicotomici descritte in letteratura utilizzate in simulazioni Monte Carlo. θ rappresenta l'abilità del soggetto in un modello IRT; v è il numero di possibili risposte all'item.

| id | nomi originali | rinomenclatura | definizione interpretata | autori |
|----|------------------|--------------------|---|--|
| 1 | | lucky guessing | $\theta < 0$. Il punteggio totale è ottenuto riassegnando probabilisticamente una percentuale di risposte incorrette ($p(X = 1) = \frac{1}{v}$). | nostra proposta |
| 2 | unlucky guessing | unlucky guessing | $\theta > 0$. Il punteggio totale è ottenuto riassegnando probabilisticamente una percentuale di risposte corrette ($p(X = 1) = \frac{1}{v}$). | St-Onge et al. (2011) |
| 3 | cheating | spuriously high | $\theta < 0$. Il punteggio totale è ottenuto cambiando un certo numero di risposte da incorrette a corrette. | St-Onge et al. (2011) |
| 4 | | spuriously low | $\theta > 0$. Il punteggio totale è ottenuto cambiando un certo numero di risposte da corrette a incorrette. | nostra proposta |
| 5 | spuriously high | spuriously high 2 | θ casuale. Il punteggio totale è ottenuto cambiando un certo numero di risposte da incorrette a corrette. | Rudner (1983), St-Onge et al. (2009) |
| 6 | spuriously low | spuriously low 2 | θ casuale. Il punteggio totale è ottenuto cambiando un certo numero di risposte da corrette a incorrette. | Rudner (1983) |
| 7 | distortion | distortion | θ casuale. Il punteggio totale è invertendo una certa percentuale di risposte. | Burnkrant e Harvey (2000) |
| 8 | lucky guessing | lucky guessing 2 | $-2 < \theta < -0.5$. Il punteggio totale è ottenuto assegnando una probabilità $p(X = 1) = \frac{1}{v}$ di risposta corretta al 41% degli item più difficili. | Karabatsos (2003) |
| 9 | cheaters | spuriously high 3 | $\theta < 0$. Il punteggio totale è ottenuto settando come corrette le risposte al 18% degli item più difficili. | Meijer et al. (1994), Meijer (1994), Karabatsos (2003) |
| 10 | careless | unlucky guessing 2 | $.5 < \theta < 2$. Il punteggio totale è ottenuto assegnando una probabilità $\frac{v}{2v}$ di risposta corretta al 41% degli item più facili. | Karabatsos (2003) |
| 11 | creative | spuriously low 3 | $.5 < \theta < 2$. Il punteggio totale è ottenuto settando come incorrette le risposte al 18% degli item più facili. | Karabatsos (2003) |
| 12 | spuriously high | spuriously high 4 | θ casuale. Il punteggio totale è ottenuto segnando come corrette una percentuale di risposte, indipendente dalla risposta non manipolata. | Levine e Rubin (1979); Drasgow (1982); Drasgow et al. (1985) |

| id | nomi originali | rinomenclatura | definizione interpretata | autori |
|----|--------------------------|----------------------|---|---|
| 13 | cheating | spuriously high 5 | θ casuale. Il punteggio totale è ottenuto segnando come corrette una certa percentuale di risposte agli item più difficili. | Emons et al. (2003) |
| 14 | spuriously low | guessing | θ casuale. Il punteggio totale è ottenuto riestraendo una certa percentuale di risposte con probabilità di risposta corretta $p(X = 1) = \frac{1}{v}$ | Levine e Rubin (1979); Drasgow (1982); Drasgow et al. (1985) |
| 15 | inattention | guessing 2 | θ casuale. Il punteggio totale è ottenuto riestraendo con una probabilità di risposta corretta $p(X = 1) = \frac{1}{v}$ una certa percentuale degli item più facili. | Emons et al. (2003) |
| 16 | Totally-random, randomly | guessing, guessing 3 | Probabilità di risposta corretta costante in ogni item $p(X = 1) = \frac{1}{v}$ | Meijer et al. (1994); Meijer (1994); Burnkrant e Harvey (2000); Karabatsos (2003) |

Sulla base delle descrizioni appena presentate sono state individuate cinque dimensioni che permettono sinteticamente di caratterizzare le singole perturbazioni delle risposte. Queste dimensioni sono:

1. **Risposta da modello.** Indica se e come la risposta nella condizione onesta influenza l'applicazione dell'aberrazione (in alcuni casi questo non avviene, in altri casi la risposta su cui applicare l'aberrazione deve essere esclusivamente 0, in altri casi esclusivamente 1).
2. **Controllo di θ .** Indica se e come il livello di abilità, θ , del soggetto scelto per la manipolazione viene controllato dallo sperimentatore (θ può essere non controllato, sopra la media, sotto la media).
3. **Effetto nel punteggio test.** Indica come il punteggio totale del test viene modificato dall'aberrazione (esso può essere aumentato, diminuito oppure, dipendentemente da θ e δ , a volte l'uno, a volte l'altro).
4. **Scelta degli item.** Indica come vengono scelti gli item da aberrare (in modo casuale, selezionando i più facili oppure i più difficili, oppure alterandoli tutti).
5. **Probabilità di manipolazione.** Indica se la risposta viene modificata deterministicamente oppure probabilisticamente dopo essere stata scelta per l'aberrazione.

Abbiamo pensato a una rinomenclatura per cercare di standardizzare i nomi delle aberrazioni: se la modifica della risposta, dopo che l'item è stato individuato come possibile item da aberrare avviene per certo, allora abbiamo usato il nome *spuriously*;

Tabella 3.2: La classificazione delle diverse aberrazioni definita dalle 5 caratteristiche identificate.

| rinomenclatura | risposta | | theta | | non cont. | diminuisce | punteggio totale | | tutti | item | | modifica | |
|-----------------------|----------|--------|-------|------|-----------|------------|------------------|---------------------|-------|---------|--------------|-----------------|-------|
| | corretta | errata | basso | alto | | | aumenta | dipende da θ | | casuali | i più facili | i più difficili | prob. |
| 1 lucky guessing | x | x | x | | | | x | | | | | | x |
| 2 unlucky guessing | | | | x | | x | | | | | | | x |
| 3 spuriously high | x | | x | | | | x | | | | | | x |
| 4 spuriously low | x | | | x | | x | | | | | | | x |
| 5 spuriously high 2 | x | | | | x | | x | | | | | | x |
| 6 spuriously low 2 | x | | | | x | | x | | | | | | x |
| 7 distortion | x | | | | x | | | x | | | | | x |
| 8 lucky guessing 2 | | NA | | | | | x | | | | | | x |
| 9 spuriously high 3 | | NA | x | | | | x | | | | | | x |
| 10 unlucky guessing 2 | | NA | | x | | x | | | | | x | | x |
| 11 spuriously low 3 | | NA | | | | | | | | | x | | x |
| 12 spuriously high 4 | | NA | | | x | | x | | | | | | x |
| 13 spuriously high 5 | | NA | | | x | | x | | | | | x | x |
| 14 guessing 2 | | NA | | | x | | | x | | | | | x |
| 15 guessing 2 | | NA | | | x | | | x | | | | | x |
| 16 guessing 3 | | NA | | | x | | NA | | x | | | | x |

se invece la modifica è probabilistica allora abbiamo usato il nome *guessing*. Nel caso *spuriously*, se l'aberrazione sposta la risposta da 0 a 1, il nome diventa *spuriously high*; se la risposta passa da 1 a 0 *spuriously low*. Nel caso *guessing*, se le probabilità sono di aumentare il punteggio si usa *lucky guessing*; se le probabilità sono di diminuire il punteggio si usa *unlucky guessing*. Infine, se aumentare o diminuire il punteggio dipende dall'abilità θ del soggetto allora la manipolazione è chiamata semplicemente *guessing*. Aberrazioni molto simili sono state aggregate. Per l'aberrazione *distortion* abbiamo tenuto il nome originale per sottolinearne la peculiarità di creare una intenzionale confusione allo sperimentatore riguardo al risultato del test. La tabella 3.2 mostra la nuova rinomenclatura e riassume le caratteristiche di ogni manipolazione rispetto alle 5 dimensioni individuate.

Al fine di sottolineare il parallelismo tra SGR e le altre metodologie di aberrazione, mostriamo come sia possibili imitare le prime sette tipologie della tabella 3.2 attraverso la metodologia SGR. Sono state scelte proprio queste perché sono già per natura simili alla metodologia SGR tradizionale poiché tengono conto della risposta data dal soggetto e ogni item ha la stessa probabilità di essere aberrato. Il primo passo è stato una scomposizione dei parametri SGR, π^+ e π^- , nei termini o e c , dove o rappresenta la proporzione di risposte da sostituire e c rappresenta la probabilità di risposta corretta nella nuova estrazione, π^+ e π^- sono due parametri dell'SGR e rappresentano, rispettivamente, la probabilità di sostituire la risposta da 0 a 1, e la probabilità di sostituire la risposta da 1 a 0.

Riportiamo le formula dell'SGR nel caso in cui gli item siano dicotomici, $v = \{0, 1\}$ (Lombardi et al., 2014) da poter confrontare con la seguente tabella 3.3:

$$p(X_f|X_d, \pi^+ \geq \pi^-) = \begin{cases} 1, & X_d = X_f = 1 \\ u, & X_d = 0, X_f = 1 \\ 1 - u, & X_d = X_f = 0 \\ 0, & X_d = 1, X_f = 0 \end{cases} \quad (3.1)$$

$$p(X_f|X_d, \pi^+ \leq \pi^-) = \begin{cases} 1 - u, & X_d = X_f = 1 \\ 0, & X_d = 0, X_f = 1 \\ 1, & X_d = X_f = 0 \\ u, & X_d = 1, X_f = 0 \end{cases} \quad (3.2)$$

dove nelle due equazioni

$$u = \max(\pi^+, \pi^-) - \min(\pi^+, \pi^-).$$

Tabella 3.3: Procedure per imitare alcune manipolazioni dicotomiche attraverso la procedura SGR. Il termine o rappresenta la proporzione di risposte da sostituire e c rappresenta la probabilità di risposta corretta nella nuova estrazione. π^+ è la proporzione di celle da sostituire con una risposta maggiore all'onesta, π^- è la proporzione di celle da sostituire con una risposta minore all'onesta.

| Aberrazione | come usare SGR per replicare l'aberrazione |
|---------------------|--|
| 1 lucky guessing | 1 generare un pattern con $\theta < 0$ 2 applicare una manipolazione SGR : $\pi^+ = oc$; $\pi^- = 0$ |
| 2 unlucky guessing | 1 generare un pattern con $\theta > 0$ 2 applicare una manipolazione SGR : $\pi^+ = 0$; $\pi^- = o(1 - c)$ |
| 3 spuriously high | 1 generare un pattern con $\theta < 0$ 2 applicare una manipolazione SGR : $\pi^+ = oc$ dove $c = 1$; $\pi^- = 0$ |
| 4 spuriously low | 1 generare un pattern con $\theta > 0$ 2 applicare una manipolazione SGR : $\pi^+ = 0$; $\pi^- = o(1 - c)$ dove $c = 0$ |
| 5 spuriously high 2 | 1 generare un pattern con θ casuale 2 applicare una manipolazione SGR : $\pi^+ = oc$ dove $c = 1$; $\pi^- = 0$ |
| 6 spuriously low 2 | 1 generare un pattern con θ casuale 2 applicare una manipolazione SGR : $\pi^+ = 0$; $\pi^- = o(1 - c)$ dove $c = 0$ |
| 7 distortion | 1 dividere i pattern in 2 gruppi: gruppo a) di numerosità $(1-o)*n$, per le risposte da non modificare ; gruppo b) di numerosità $o*n$, per le risposte da manipolare 2 Nel gruppo di risposte da distorcere (b) si applica il modello di sostituzione: $p(X_f X_d) = \begin{cases} c, & X_d \neq X_f \\ 1 - c, & X_d = X_f \end{cases}$ dove c , in questo caso, rappresenta la probabilità di inversione |

3.2.2 Codifica politomica ordinale della risposta

In letteratura non sono presenti molti lavori che utilizzino aberrazioni su risposte ordinali, come ad esempio scale Likert, nonostante i test di questo tipo siano molto diffusi e spesso utilizzati in vari ambiti di ricerca psicologica.

Glas e Dagohey (2007, pg. 167) utilizzano un test di personalità le cui possibili risposte agli item vengono date su una scala Likert a 5 punti. I ricercatori manipolano

i parametri del modello generativo, in particolare aumentano l'abilità dei soggetti di un fattore pari a 1 quando calcolano le probabilità di risposta per una parte del test e chiamano questa manipolazione *different abilities for different subsets of items*. Utilizzano anche la manipolazione *guessing on part of the items* dove per una sezione del test viene fissata a $\frac{1}{v}$ la probabilità di ogni livello della scala.

Emons (2008, pg. 229-230) utilizza le tre seguenti aberrazioni *Carelessness and inattention*, dove simula un comportamento di risposta casuale ed ogni risposta ha la stessa probabilità di essere scelta $p(X_f) = \frac{1}{v}$. La seconda manipolazione è chiamata *Tendency to choose extreme response options*, e indica la tendenza del soggetto a scegliere le risposte agli estremi della scala. Per simulare tale comportamento gli autori traslano i parametri generativi delle soglie, spostandoli verso il centro, generando di conseguenza un aumento delle probabilità per le risposte estreme. Infine il Reversed scoring consiste nel ricodificare le risposte, precedentemente generate, invertendone il valore, $X_f = v - X_d$.

È importante menzionare anche altri autori che hanno trattato la simulazione del faking. Ad esempio, Ferrando e Anguiano-Carrasco (2013) attribuiscono a Zickar e Robie (1999) un modello per formalizzare il fenomeno del faking: un meccanismo chiamato *theta-shift*, il quale assume che il tratto individuale (θ) che si sta misurando sia momentaneamente aumentato nel soggetto quando è presente la condizione di faking, col risultato di incrementare il punteggio atteso totale del test che coincide, praticamente allo stesso meccanismo introdotto da Glas e Dagohoy (2007).

3.3 Studio simulativo

Scopo di questa sezione è esplorare come una selezione di *person fit statistics* (PFS): Number of Guttman Errors G^p , Normed Number of Guttman Errors G_N^p , Generalized U3 Person-Fit Statistic $U3^p$; Emons (2008); Parametric Polytomous Person-Fit Statistic $I_{\frac{p}{2}}^p$; Drasgow et al. (1985), riescano a discriminare tra pattern aberrati e non, nella situazione in cui le aberrazioni siano introdotte attraverso i modelli di faking Sample Generation by Replacement, SGR (Lombardi et al., 2014).

Si sono scelti di indagare gli indici politomici perché l'utilizzo e gli effetti delle perturbazioni nel caso dicotomico sono già stati largamente indagati da altri autori, mentre per il caso politomico la letteratura è carente sia per modellistica di perturbazione sia per analisi delle performance degli indici PFS, inoltre riteniamo che il caso dicotomico rappresenti solo una semplificazione del caso più complesso associato alla modellizzazione politomica ordinale.

Questa sezione è organizzata come segue: nella prima parte verranno illustrati i quattro indici PFS utilizzati in questo studio, a seguire viene esplicitata nel dettaglio la metodologia SGR, cioè sono definiti il modello generativo e i modelli perturbativi dei dati, la terza parte descrive il disegno sperimentale della simulazione MC e infine sono riportati i risultati della simulazione.

3.3.1 Quattro indici di fit del soggetto

I Person-Fit Statistic sono degli indici statistici che cercano di discriminare pattern di risposta aberranti dai pattern coerenti. Essi sono stati originariamente costruiti per identificare aberrazioni del pattern di risposta come il copiare alcune risposte corrette, oppure rispondere a caso (Ziegler et al., 2012). Ci teniamo a ribadire, invece, che il faking è un comportamento intenzionale, atto ad alterare le risposte con l'obiettivo di dichiarare un livello di costrutto diverso da quello che si dichiarerebbe in assenza di faking, non in test di abilità, ma in scale self-report.

Una distinzione importante che si può fare per i PFS politomici è quella tra indici basati su modelli teorici (model based index) e indici che non si basano su un modello ma solo sui dati osservati (non-parametric index). Nel primo caso l'indice misura la verosimiglianza di un pattern di risposte rispetto al livello di costrutto latente calcolato per quel soggetto, per esempio l_z^p , (Drasgow et al., 1985). Il secondo tipo di indici si basa esclusivamente sui dati raccolti, stima una coerenza tra le risposte, per esempio G^p , (Emons, 2008).

Per una trattazione teorica degli indici politomici non parametrici che verranno utilizzati in questa simulazione (G^p , G_N^p , $U3^p$) si consiglia Emons (2008), per l_z^p invece Drasgow et al. (1985). Per un approfondimento sui numerosi indici dicotomici si consiglia la review di Meijer e Sijtsma (2001).

l_z politomico

L'indice l_z^p valuta la verosimiglianza di un determinato pattern di risposte moltiplicando le probabilità di ogni risposta data da un soggetto il quale possiede un determinato livello di costrutto stimato. L'indice l_z^p è una misura standardizzata secondo una distribuzione normale che permette di identificare i soggetti che si collocano sulla coda sinistra della distribuzione che sono caratterizzati da profili di punteggi anomali o improbabili secondo il modello.

Number of Guttman Errors

Questo indice calcola l'incoerenza tra le risposte misurando la "distanza" in soglie tra tutte le risposte date confrontando tutte le coppie possibili di item. Dopo aver unito le soglie di due item in un nuovo vettore, le si ordina in base alla loro difficoltà. Si conta la distanza in soglie tra le risposte date ai due item. Questa operazione si ripete per tutte le coppie di item e si sommano i punteggi, tale somma è denominata G^p . Il range va da 0 a T (dove T è un valore che dipende dalle caratteristiche strutturali del test). Il valore 0 indica perfetta coerenza tra le risposte, al contrario un valore alto è indice di maggiore incoerenza.

Normed Number of Guttman Errors

La statistica G_N^p è la normalizzazione di G^p . Dato che G_N^p dipende dalla numerosità degli item e dal numero di livelli di ogni item e dal punteggio test, è lecito dividere G^p per il numero massimo di possibili di errori di Guttman per uno specifico punteggio del test osservato. Il range va da 0 a 1: 0 significa perfetta coerenza tra risposte, 1 massima incoerenza.

Generalized U3 Person-Fit Statistic

$U3^p$ è un indice di tipo non parametrico che assomiglia a l_z^p , ma invece di utilizzare la probabilità di dare un certa risposta calcolata da modello, utilizza una procedura non-parametrica per stimare la percentuale di soggetti che hanno passato la soglia che è necessaria per produrre una risposta corretta. Questo indice valuta l'improbabilità di ottenere un particolare punteggio test con esattamente le risposte fornite dal soggetto. E' un indice normalizzato e il cui range va da 0 a 1: 0 significa perfetta coerenza tra risposte, 1 massima incoerenza.

3.3.2 Modello di risposta onesta

Il modello scelto per la generazione dei dati sotto la condizione onesta (non perturbata) è il modello IRT: Generalized Partial Credit Model (GPCM Muraki, 1993). Abbiamo scelto questo particolare modello IRT perché in un ottica simulativa di tipo MC, l'*Item Response Function* (IRF) è direttamente controllabile regolando i parametri dei confini tra ogni categoria e ci risultava più semplice creare degli item con delle buone caratteristiche psicometriche (sufficiente discriminatività delle categorie e rispetto dell'ordinamento delle stesse).

3.3.3 Modelli di manipolazione fake

La seconda parte della metodologia SGR consiste nel manipolare i dati precedentemente generati (utilizzando il modello GPCM) attraverso un modello di sostituzione, in questo caso \mathbf{Z}_x . Nel nostro studio di simulazione descritto in questo capitolo limiteremo l'ambito di analisi ai due scenari più importanti di faking good: a) faking good slight b) faking good extreme. Il faking good può essere definito come un tentativo cosciente da parte dell'individuo di presentare informazioni false per creare una impressione più positiva con l'obiettivo di influenzare altre persone (es., Furnham, 1986; Zickar & Robie, 1999; McFarland & Ryan, 2000). In generale, i manipolatori fake-good modificano i loro punteggi sulla scala in modo da fornire punteggi più estremi (es., Furnham, 1986; Viswesvaran & Ones, 1999; McFarland & Ryan, 2000; Griffin et al., 2004). Ovviamente la direzione dello spostamento dipende dal significato dell'item, ma qui ipotizziamo gli item siano omogeneamente di tipo monodirezionali. Una manipolazione fake-good rappresenta una situazione dove le risposte sono soggette esclusivamente a uno spostamento positivo:

$$X^f \geq X^d.$$

Slight ed extreme si differenziano principalmente per l'intensità della manipolazione, per esempio, una manipolazione faking good slight potrebbe essere associata a situazioni di selezione in cui il soggetto cerca di aumentare le probabilità di essere assunto ma è limitato dal rischio di invalidare il questionario facendosi notare come mentitore. Una manipolazione faking good extreme potrebbe accadere quando, fraudolentemente e sfacciatamente, si cerca di far salire la valutazione online di una propria attività commerciale pubblicizzata sulla rete.

Nell'approccio SGR il meccanismo di faking è descritto dal modello di sostituzione (Replacing Model) \mathbf{Z}_x , il quale è parametrizzato secondo una versione discreta della distribuzione beta generalizzata, con parametri $\theta_{fx} = (\gamma_x, \delta_x, \pi_x)$ (vedasi il primo capitolo e Lombardi & Pastore, 2012; Lombardi et al., 2014). Questa parametrizzazione del modello è molto flessibile e può facilmente caratterizzare le due tipologie (good e slight) di faking con diversi livelli di intensità settando adeguatamente i parametri γ e δ . Per ulteriori dettagli, il lettore può fare riferimento ai lavori originali che introducono il modello di parametrizzazione nell'impianto SGR (es., Lombardi & Pastore, 2012; Lombardi et al., 2014).

3.3.4 Il disegno sperimentale

Prendendo spunto dalla simulazione di Meijer, Niessen, e Tendeiro (2015), abbiamo analizzato la percentuale di veri positivi fissata una soglia, utilizzando i quattro indici

di aberrazione politomici G^p , G_N^p , $U3^p$, l_z^p . Per l_z^p abbiamo indagato i soggetti con un punteggio minore del 5° percentile dei punteggi del campione, mentre per gli altri indici abbiamo utilizzato il 95° (nota: un punteggio di un'aberrazione dovrebbe essere probabilisticamente minore rispetto a un punteggio onesto se si usa l_z^p , viceversa per gli altri indici). Sei fattori sono stati fatti variare sistematicamente in un disegno sperimentale completo, per ognuno di essi è specificata la motivazione che ci ha guidati alla scelta dello stesso e dei suoi livelli:

- numerosità degli item a 3 livelli: $J = 5, 10, 20$;
Quando aumenta la lunghezza del test, aumenta anche la percentuale aberrazioni rilevate (St-Onge et al., 2009);
- numero di livelli di risposta a 2 livelli: $V = 5, 7$;
Glas e Dagohoy (2007, pg. 168) utilizzano una scala a 5 livelli; Emons (2008) analizza due scale, una a 3 e una a 5.
- numerosità dei soggetti, a 3 livelli: $N = 300, 600, 900$;
Il tallone di Achille dei PFS è che necessitano di un gran numero di soggetti per produrre delle stime attendibili dei parametri, ma spesso la numerosità N è limitata (Emons et al., 2003) quindi utilizziamo numerosità campionarie sufficientemente grandi;
- percentuali di pattern aberranti, a 3 livelli: $A = 5\%, 10\%, 25\%$;
Solitamente, nella letteratura consultata, la percentuale di profili fake è quasi sempre bassa 5-10%, oppure il modello viene fittato in un dataset onesto e poi identificati i fakers nel campione preso in esame, però abbiamo ritenuto questa seconda pratica irrealizzabile in una situazione non simulata dove non si hanno soggetti in grandi quantità. Karabatsos (2003) utilizza le seguenti percentuali di pattern aberrati: 5%, 10%, 25%, 50%;
- livelli di intensità dell'aberrazione, a 4 livelli: $P = .25, .50, .75, .9$;
Riguardo l'intensità dell'aberrazione (probabilità di ogni risposta di essere aberrata in un pattern menzognero da una manipolazione faking good) il più grande range oggetto di studio l'abbiamo trovato in St-Onge et al. (2011, pg. 420), dove per analizzare la relazione tra percentuale di risposte aberrate e l'accuratezza dei PFS, gli autori fanno variare la percentuale di risposta aberrate tra il 10% e il 60% con un incremento dello 2,5%. Loro scelgono questo intervallo perché: a) conteneva le probabilità di aberrazione già utilizzate in precedenza, permettendone il confronto, e b) percentuali più alte di aberrazioni permettevano di esplorare nuovi range (precedentemente era solo fino al 41%);

- tipi di modello di faking good, a 3 livelli: $M=extreme, slight, random$.

A differenza degli altri due modelli di sostituzione, *random* è bidirezionale e uniforme all'interno delle direzioni, $p(X_f < X_d) = p(X_d < X_f) = \pi/2$. Tale perturbazione è definita dalla probabilità condizionale:

$$p(X_f|X_d) = \begin{cases} \frac{\pi}{2(X_d-1)}, & \forall X_f \in 1 \leq X_f < X_d < v \\ 1 - \pi, & \exists! X_f = X_d \notin 1 \cup v \\ 1 - \pi/2, & \exists! X_f = X_d \in 1 \cup v \\ \frac{\pi}{2(v-X_d)}, & \forall X_f \in 1 < X_d < X_f \leq v \end{cases} \quad (3.3)$$

Circa il modello generativo iniziale (modello onesto), St-Onge et al. (2009) fanno notare che si può ipotizzare che i modelli di risposta inappropriati siano più facili da individuare quando i test comprendono elementi (item) altamente discriminanti con diversi livelli di difficoltà, riformulando: un test discrimina meglio i pattern aberranti se ha buone proprietà psicometriche, quindi abbiamo stabilito la modalità di generazione dei parametri del modello IRT, i quali verranno riestratti ad ogni ripetizione. Il modello era un GPCM (Muraki, 1993) utilizzato nella nostra simulazione è costituito dai seguenti parametri θ_j :

- θ_s : le abilità di ogni soggetto s , estratte da una distribuzione normale standardizzata;
- a_j : il parametro di discriminazione, estratto da una distribuzione log-normale, $\mu = 0$, $\sigma = .25$, come in Glas e Dagohey (2007);
- d_{jc} , le soglie della categorie: saranno definiti a partire da questi valori arbitrari in funzione di V :

| | | | | | | | |
|---------------------|-----------------|------|-----|-----|----|---|-----|
| $V = 5 \rightarrow$ | <i>soglie</i> : | -1 | -.5 | .5 | 1 | | |
| $V = 7 \rightarrow$ | <i>soglie</i> : | -1.5 | -1 | -.5 | .5 | 1 | 1.5 |

 aggiungereemo inoltre, ad ogni soglia, un errore casuale preso da una distribuzione normale troncata ($\sigma = .1519892$) per aumentare la variabilità.¹
- b_j : il parametro di traslazione, sarà estratto da una normale standardizzata e rimarrà costante all'interno dell'item.

¹Si è calcolato questo valore di σ , facendo riferimento al caso con $V = 7$ dove le soglie sono più vicine tra loro, accettando che nel 5% dei casi possa capitare che i valori assegnati a un parametro possano superare il punto mediano tra due soglie adiacenti. Si è poi deciso di troncare la normale proprio al punto medio tra due soglie per evitare che una soglia teoricamente più bassa superi una teoricamente superiore, e viceversa. In pratica si è cercato un compromesso tra aggiungere variabilità e non creare inutile errore generando item di bassa qualità le cui soglie si sovrapponevano.

Definiamo con j, n, v, α, π e m i distinti livelli dei fattori J, N, V, A, P e M , rispettivamente. I seguenti passaggi procedurali sono stati ripetuti 1,000 volte per ognuna delle $648 = 3 \times 2 \times 3 \times 3 \times 4 \times 3$ combinazioni dei livelli dei fattori (J, N, V, A, P e M) del disegno di simulazione:

1. Generazione di un dataset di dati grezzi \mathbf{D} con dimensione n secondo i parametri IRT onesti θ_j . Più specificatamente, ogni cella di \mathbf{D} è campionata secondo il modello GPCM combinando il parametro di abilità del soggetto θ_s con i parametri che caratterizzano l'item (a_j, d_{jc}, b_j)
2. Per ogni osservazione $s = 1, \dots, n$ viene estratta una variabile a dicotomica (0/1) sulla base della proporzione di fakers A . Se $g = 1$, allora la g -esima riga di \mathbf{D} viene sostituita da una nuova riga, riga di sostituzione, ottenuta sulla base del modello di faking M . Al contrario, se $g = 0$, viene mantenuta la riga s -th di \mathbf{D} . Questa procedura porta alla costruzione di una nuova matrice mista \mathbf{X} .
3. Viene calcolata e salvata la percentuale di fakers identificati come tali nel 5% più estremo per ogni indice $G^p, G_N^p, U3^p, l_z^p$.

L'intera procedura ha generato un totale di $6,480 = 1,000 \times 3 \times 2 \times 3 \times 3 \times 4 \times 3$ matrici corrotte, matrici non corrotte e anche quartetti di percentuali di veri fakers identificati come tali². L'appendice A riporta alcuni dettagli sugli aspetti di ottimizzazione basati sul calcolo parallelo che sono stati implementati in questa simulazione MC.

3.3.5 Analisi dei dati

Attraverso un'analisi grafica abbiamo descritto il comportamento degli indici al variare dei livelli dei fattori sperimentali. Nelle Figure dalla 3.1 alla 3.4 sono riportati valori della mediana della percentuale degli hit (fakers riconosciuti come tali) e gli intervalli di confidenza al 2.5° e 97.5° percentile. I grafici così disegnati sono molto informativi perché dallo stesso grafico si possono estrarre tutte e quattro le celle della tabella di classificazione (hit, miss, falsi positivi e veri negativi). Per confrontare le prestazioni degli indici, fissata la percentuale di fakers (rappresentata nelle colonne), più alta è la percentuale ottenuta, migliore è la prestazione. Un esempio concreto: riferendoci alla Figura 3.1, nel caso di una percentuale di fakers del 25% che manipolano il 25% delle risposte secondo un modello slight (seconda riga, terza colonna), l'indice l_z^p identifica

²Nei passaggi 1 e 3, il campione generato, rispettivamente \mathbf{D} oppure \mathbf{X} , è stato ricampionato nel caso in cui una delle due matrici non avesse almeno una frequenza per ogni livello di ogni item, ciò è stato fatto per non incorrere in problemi di calcolo degli indici PFS.

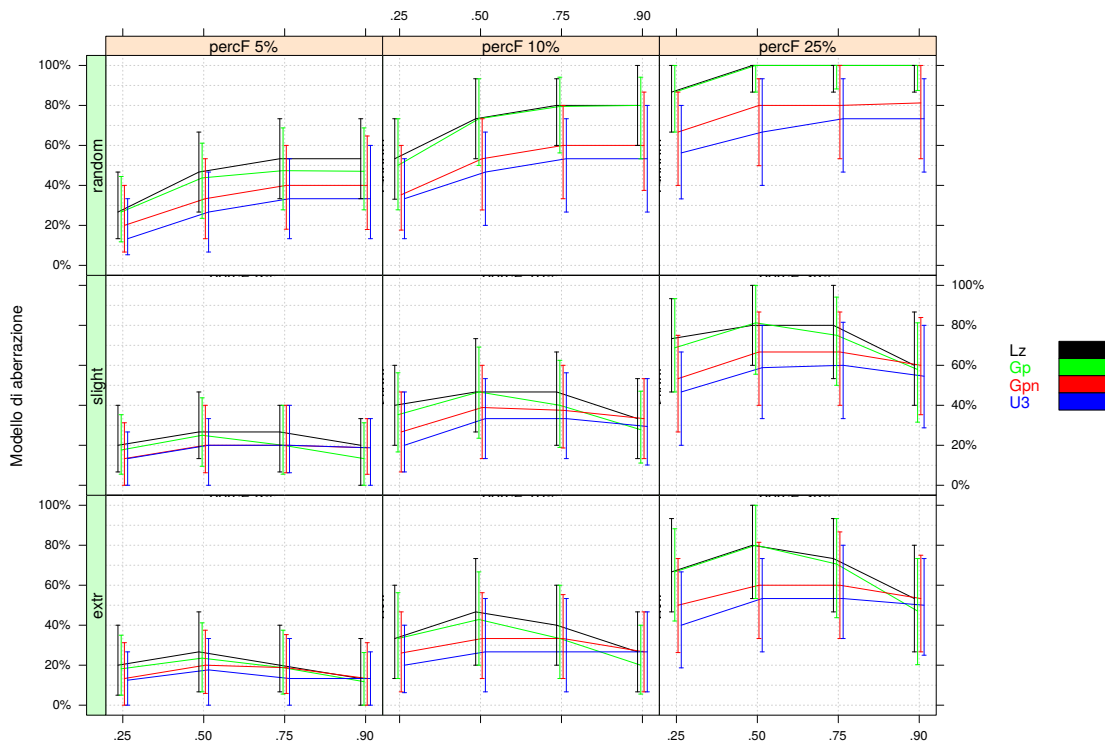


Figura 3.1: Percentuale di veri positivi sul totale di soggetti segnalati come funzione della probabilità di aberrazione di una risposta, il modello di faking e la percentuale di fakers nel campione, tenendo costanti il numero di livelli dell'item (5), il numero di item (5) e la dimensione campionaria (300).

correttamente circa il 75% dei soggetti che segnala come potenziali manipolatori, mentre $U3^p$ riesce ad identificarne correttamente il 50%. La percentuale mancante (25% nel primo caso e 50% nel secondo) sono soggetti onesti, ma segnalati come potenziali fakers; in questo caso sarebbe da preferire l'utilizzo di l_z^p .

Riportiamo solo i grafici con più bassa numerosità campionaria $N = 300$ perché anche aumentando il campione non si notano differenze sostanziali. Riportiamo solo $J = 5$ e 20 (numerosità degli item), i risultati per 10 item, sono intermedi.

Salvo pochi casi, l'indice più performante sembra essere l_z^p , spesso equiparato dall'indice G^p (il migliore dei non parametrici), a seguire G_N^p e $U3^p$. Questo ordine sembra essere abbastanza stabile attraverso tutti i fattori sperimentali. C'è un'interazione tra il modello di faking, M , e la probabilità di aberrazione di una risposta, P : i pattern manipolati da faking (slight ed extreme) sono più facilmente identificabili quando la percentuale di risposte aberrate è circa il 50%, ma diminuisce la probabilità di identificazione all'aumentare delle risposte manipolate, ciò forma una relazione di tipo concavo. Il modello di faking extreme ha una percentuale di identificazione paragonabile al faking slight. Invece, per il modello di manipolazione random l'identificazione

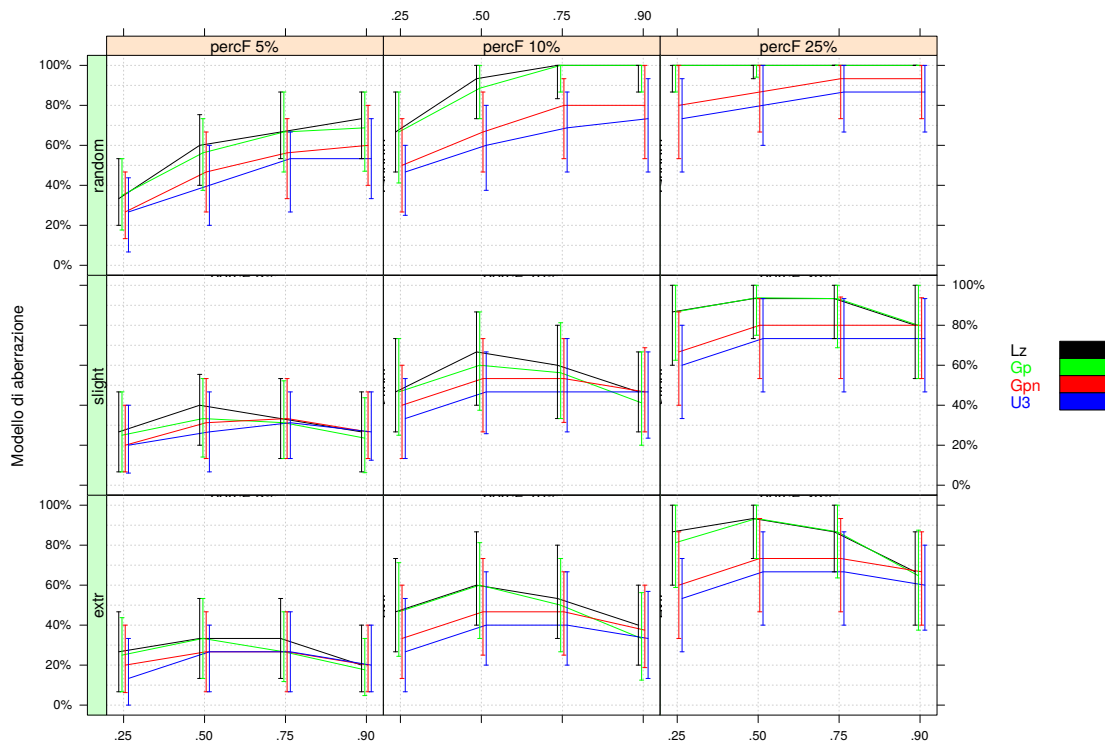


Figura 3.2: Percentuale di veri positivi sul totale di soggetti segnalati come funzione della probabilità di aberrazione di una risposta, il modello di faking e la percentuale di fakers nel campione, tenendo costanti il numero di livelli dell'item (7), il numero di item (5) e la dimensione campionaria (300).

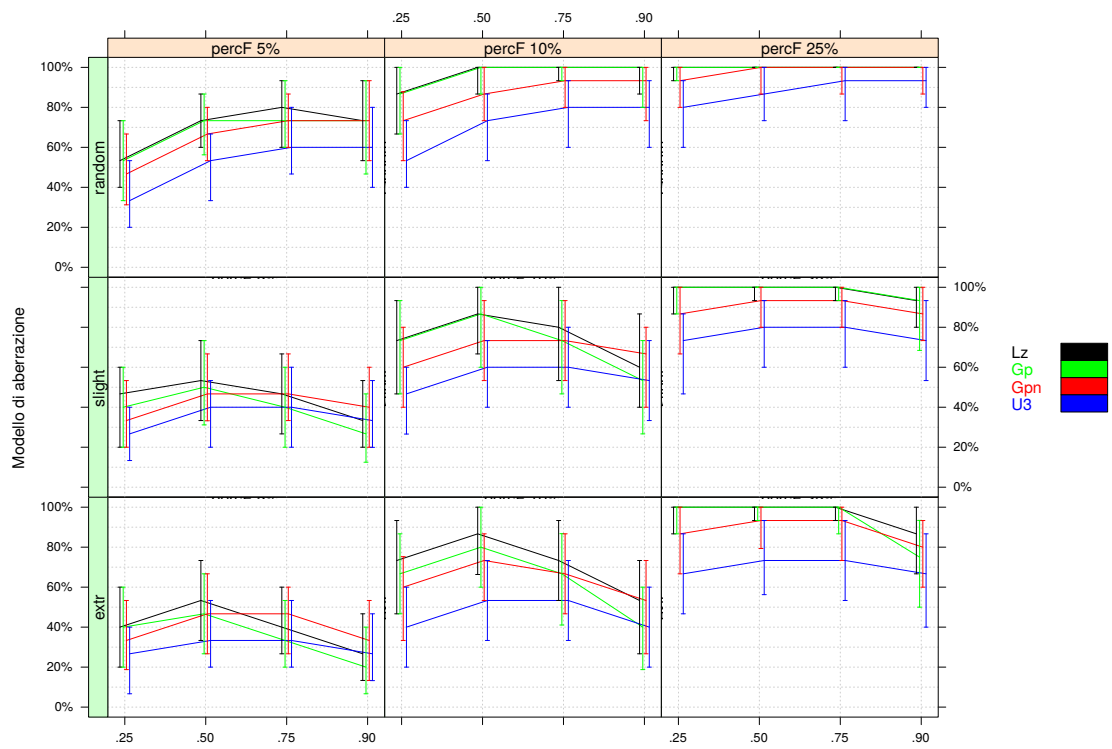


Figura 3.3: Percentuale di veri positivi sul totale di soggetti segnalati come funzione della probabilità di aberrazione di una risposta, il modello di faking e la percentuale di fakers nel campione, tenendo costanti il numero di livelli dell'item (5), il numero di item (20) e la dimensione campionaria (300).

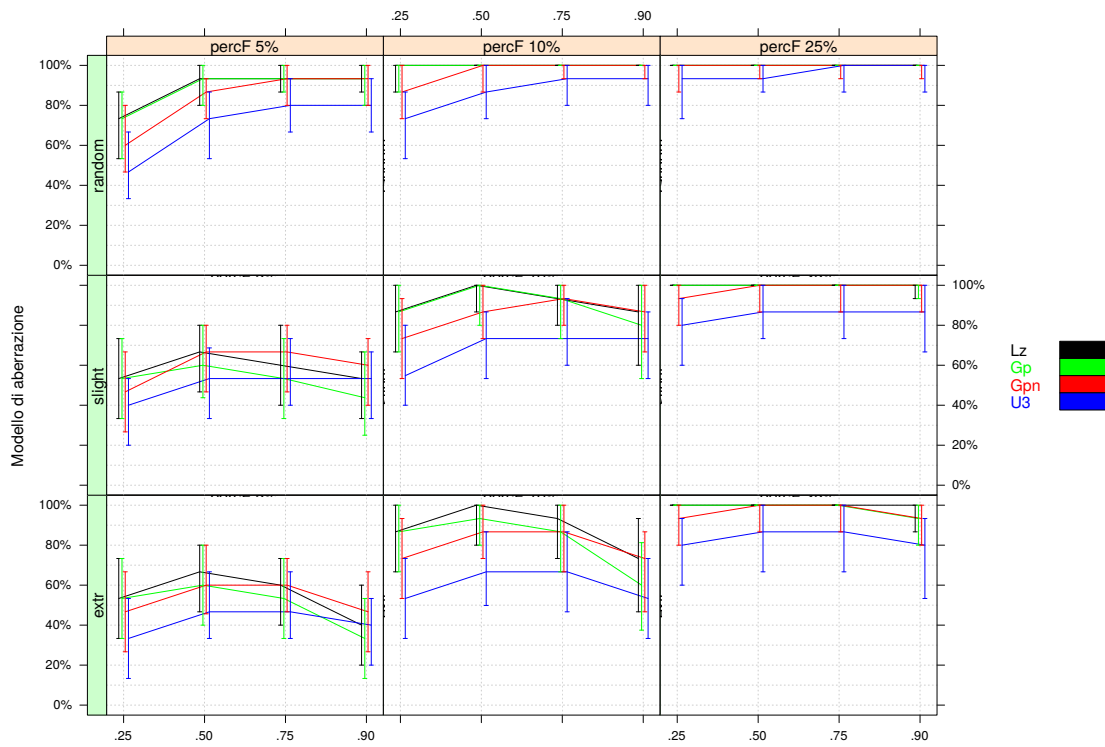


Figura 3.4: Percentuale di veri positivi sul totale di soggetti segnalati come funzione della probabilità di aberrazione di una risposta, il modello di faking e la percentuale di fakers nel campione, tenendo costanti il numero di livelli dell'item (7), il numero di item (20) e la dimensione campionaria (300).

corretta è mediamente superiore agli altri modelli ed aumenta all'aumentare di A .

L'incremento di numerosità dei livelli di risposta, V , aumenta la corretta identificazione dei profili manipolati. Aumentare il numero degli item J aumenta sensibilmente la prestazione degli indici.

3.4 Analisi esplorativa della Likelihood dei pattern e dataset soggetti a faking

Nella precedente sezione, abbiamo indagato la prestazione dei Person Fit Statistics (PFS) in dataset aberrati da Faking. Abbiamo osservato che i PFS sono poco potenti se utilizzati direttamente sulla matrice \mathbf{X} (la matrice che contiene sia dati onesti che una percentuale di dati manipolati da faking). Inoltre gli indici non danno indicazioni precise su quale sia la percentuale di fakers e sta allo sperimentatore scegliere a priori che percentuale di pattern infrequenti (improbabili, incoerenti) è necessario eliminare dal campione.

In questa nuova simulazione, che possiamo definire anche "caso critico" per i pochi fattori sperimentali manipolati, utilizzeremo una generazione dei dati multivariata, UVA, partendo dalle soglie (thresholds) e dalla matrice delle correlazioni degli item (Muthén, 1984; Jöreskog & Sörbom, 2001). Abbiamo scelto di cambiare modello generativo per sottolineare l'indipendenza della metodologia SGR da un determinato modello generativo onesto anche con un esempio pratico (nella precedente simulazione si era usato un modello IRT, par. 3.3.4).

In questo capitolo abbiamo scelto di focalizzare l'attenzione su l_z^p in quanto risultato essere l'indice più performante nella simulazione precedente. Siccome il calcolo di l_z^p per un soggetto si può riassumere come una normalizzazione del punteggio della sua Likelihood, l_z^p è una trasformazione lineare della Likelihood calcolata sul soggetto quindi, se esiste una differenza tangibile tra soggetti onesti e soggetti fakers, questa differenza si dovrebbe osservare sia in l_z^p che nella Likelihood del soggetto.

In questo capitolo la Likelihood dei soggetti verrà chiamata Likelihood locale \mathcal{L}_l per contraddistinguerla dalla Likelihood calcolata sull'intera matrice, Likelihood globale \mathcal{L}_g . Attraverso un'analisi grafica delle distribuzioni, abbiamo osservato il comportamento delle \mathcal{L}_g (della matrice) e \mathcal{L}_l (di soggetti), in matrici \mathbf{D} (matrici completamente oneste), \mathbf{F} (matrici completamente corrotte da faking) e \mathbf{X} (matrici parzialmente corrotte da faking), con gli obiettivi di indagare: a) se esiste una tangibile differenza tra la \mathcal{L}_l di un soggetto onesto e quella di un soggetto fakers b) se esiste una tangibile differenza tra la \mathcal{L}_g di una matrice di dati completamente onesta e quella di un ma matrice

contenente soggetti menzogneri.³ Se ci fosse una differenza tra le \mathcal{L}_l dei soggetti onesti e fakers, allora sarebbe possibile identificare i fakers. Se, invece, a livello di matrice si osservasse una qualche regolarità nella \mathcal{L}_g in matrici sempre più manipolate, allora si potrebbe fare un'ipotesi sulla proporzione di soggetti fakers presenti nel campione.

Per rendere più comprensibili i grafici, le \mathcal{L} sono state trasformate nel loro rispettivo logaritmico. $\log(\mathcal{L}_l)$ e $\log(\mathcal{L}_g)$ vengono definite come:

$$\log(\mathcal{L}_l) = \log \prod_j^{12} p(x_{sj} | \boldsymbol{\theta}_D); \quad (3.4)$$

$$\log(\mathcal{L}_g) = \log \prod_s^{200} \prod_j^{12} p(x_{sj} | \boldsymbol{\theta}_D); \quad (3.5)$$

dove $j = \{1, \dots, 12\}$ rappresenta l'indicatore per gli item e $s = \{1, \dots, 200\}$ rappresenta l'indicatore dei soggetti.

Abbiamo utilizzato i seguenti parametri $\boldsymbol{\theta}_D$ (tabelle 3.4 e 3.5) per la generazione di \mathbf{D} , ricavati da dati empirici, presi dalla prima somministrazione del questionario Perceived Empathic Self-Efficacy Scale, Adult version (AEP/A, Caprara, 2001) utilizzato nel lavoro di Pastore, Nucci, Bobbio, e Lombardi (2017). I parametri si basano su una somministrazione in cui veniva chiesto ai partecipanti di essere onesti riguardo alle risposte; alla somministrazione parteciparono 378 studenti della Scuola di Psicologia dell'Università di Padova. Nella tabella 3.4 sono riportate le soglie tra le categorie degli item, nella tabella 3.5 sono riportate le correlazioni tra gli item. Come si può dedurre dai parametri, gli item sono 12 e il numero di livelli di risposta per ogni item è pari a 5.

³L'analisi della \mathcal{L} è test-specifica dato che qualsiasi cambiamento ai parametri degli item andrebbe ad impattare sulle L locali; e la numerosità campionaria, oltre ai parametri degli item, sulla \mathcal{L} globale.

Tabella 3.4: Soglie dei 12 item del questionario AEP/A

| Item | Soglie | | | | | |
|------|--------|-------|-------|-------|------|------|
| | 1 | -Inf | -3.37 | -1.55 | 0.16 | 2.17 |
| 2 | -Inf | -3.37 | -1.62 | -0.63 | 1.04 | +Inf |
| 3 | -Inf | -2.00 | -0.52 | 0.73 | 1.69 | +Inf |
| 4 | -Inf | -3.37 | -1.77 | -0.30 | 1.04 | +Inf |
| 5 | -Inf | -3.37 | -2.16 | -0.78 | 0.97 | +Inf |
| 6 | -Inf | -2.43 | -1.10 | 0.26 | 1.44 | +Inf |
| 7 | -Inf | -2.17 | -0.70 | 0.27 | 1.78 | +Inf |
| 8 | -Inf | -2.43 | -1.55 | -0.54 | 0.70 | +Inf |
| 9 | -Inf | -2.43 | -0.80 | 0.36 | 1.62 | +Inf |
| 10 | -Inf | -3.37 | -2.00 | -0.54 | 0.89 | +Inf |
| 11 | -Inf | -2.43 | -2.17 | -0.52 | 0.78 | +Inf |
| 11 | -Inf | -2.43 | -0.78 | 0.36 | 1.39 | +Inf |

Tabella 3.5: Correlazione tra i 12 item del questionario AEP/A

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|-----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|
| 1 | 1.00 | 0.12 | 0.23 | 0.25 | 0.14 | 0.08 | 0.20 | 0.11 | 0.29 | 0.14 | 0.24 | 0.19 |
| 2 | 0.12 | 1.00 | 0.35 | 0.20 | 0.49 | 0.20 | 0.28 | 0.47 | 0.32 | 0.27 | 0.35 | 0.21 |
| 3 | 0.23 | 0.35 | 1.00 | 0.43 | 0.25 | 0.03 | 0.20 | 0.28 | 0.47 | 0.16 | 0.40 | 0.10 |
| 4 | 0.25 | 0.20 | 0.43 | 1.00 | 0.18 | 0.15 | 0.20 | 0.01 | 0.37 | 0.20 | 0.12 | 0.02 |
| 5 | 0.14 | 0.49 | 0.25 | 0.18 | 1.00 | 0.39 | 0.26 | 0.29 | 0.14 | 0.31 | 0.37 | 0.26 |
| 6 | 0.08 | 0.20 | 0.03 | 0.15 | 0.39 | 1.00 | 0.14 | 0.01 | 0.01 | 0.29 | 0.24 | 0.30 |
| 7 | 0.20 | 0.28 | 0.20 | 0.20 | 0.26 | 0.14 | 1.00 | 0.17 | 0.25 | 0.15 | 0.24 | 0.34 |
| 8 | 0.11 | 0.47 | 0.28 | 0.01 | 0.29 | 0.01 | 0.17 | 1.00 | 0.29 | 0.50 | 0.48 | 0.29 |
| 9 | 0.29 | 0.32 | 0.47 | 0.37 | 0.14 | 0.01 | 0.25 | 0.29 | 1.00 | 0.00 | 0.33 | 0.31 |
| 10 | 0.14 | 0.27 | 0.16 | 0.20 | 0.31 | 0.29 | 0.15 | 0.50 | 0.00 | 1.00 | 0.54 | 0.32 |
| 11 | 0.24 | 0.35 | 0.40 | 0.12 | 0.37 | 0.24 | 0.24 | 0.48 | 0.33 | 0.54 | 1.00 | 0.38 |
| 12 | 0.19 | 0.21 | 0.10 | 0.02 | 0.26 | 0.30 | 0.34 | 0.29 | 0.31 | 0.32 | 0.38 | 1.00 |

3.4.1 Analisi Likelihood locale

Per simulare un caso limite di faking, forse il più pericoloso per impatto nei dati e il più remunerativo per i soggetti fakers, si è scelto il modello di faking extreme, dove il soggetto menzionario mente il più possibile e dà, tendenzialmente, una risposta verso il fondo della scala. Questo tipo di soggetti dovrebbero essere i più importanti

da identificare dato che hanno sia un grande impatto nel proprio ranking rispetto al gruppo, sia impattano modificando le relazioni tra gli item.

Sono state generate 100 matrici \mathbf{D} secondo il modello non corrotto per osservare la distribuzione delle $\log \mathcal{L}_l$; 100 matrici \mathbf{F} completamente manipolate ($\pi = .9$, faking di tipo extreme); 100 matrici \mathbf{R} completamente con risposte casuali, $p(X_f = v) = \frac{1}{5}$, dove $v = \{1, \dots, 5\}$. Ogni matrice contiene 200 soggetti.

Nella Figura 3.5 si possono osservare le distribuzioni delle $\log \mathcal{L}_l$ per ogni matrice \mathbf{D} , \mathbf{F} , \mathbf{R} . Si può osservare come le distribuzioni della $\log \mathcal{L}_l$ nelle matrici \mathbf{D} abbiano una certa regolarità, in nero. Se queste distribuzioni vengono confrontate con delle risposte completamente casuali, contenute nelle matrici \mathbf{R} in verde, si può notare una sostanziale differenza e si può accettare verosimilmente di costruire un cut-off per discriminare i due tipi di soggetti (in questo esempio, graficamente, -19). Mentre per le distribuzioni di pattern completamente manipolata da faking, contenuti nelle matrici \mathbf{F} in rosso, è difficile pensare di poter discriminare i due tipi di soggetti, utilizzando la L_l , senza accettare grossolani margini di errore.

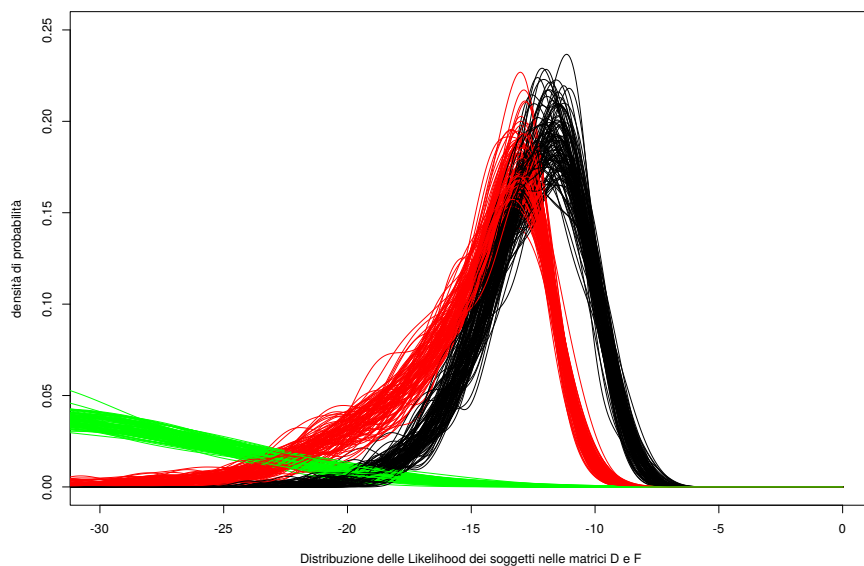


Figura 3.5: Distribuzioni delle $\log \mathcal{L}_l$ di pattern generati da modello, in nero, completamente aberrati, in rosso, e casuali, in verde

3.4.2 Analisi Likelihood globale

Dopo aver analizzato la $\log \mathcal{L}_l$ e aver concluso che non è un indice sufficientemente sensibile al faking, ci siamo dedicati all'analisi della $\log \mathcal{L}_g$ (della matrice). Nella Figura 3.6, si possono osservare la distribuzione di 1000 matrici da 200 soggetti generate secondo modello, in nero, la distribuzione dello stesso numero di matrici completamente

aberrate da faking 1000 matrici, in rosso. In grigio sono rappresentate altre distribuzioni di $\log \mathcal{L}_g$ di matrici con una percentuale di fakers variabile, dal 10%, 20%, ..., 90%, 1000 matrici ognuna. Da destra, verso sinistra, la $\log \mathcal{L}_g$ diminuisce all'aumentare della percentuale di fakers. In questo caso le $\log \mathcal{L}_g$ sembrano essere più affidabili tanto da, noto un modello generativo non corrotto, scelto un modello di faking e osservata una certa $\log \mathcal{L}_g$ di una matrice mista \mathbf{X} , poter ipotizzare la percentuale più probabile di soggetti fakers nella matrice \mathbf{X} .

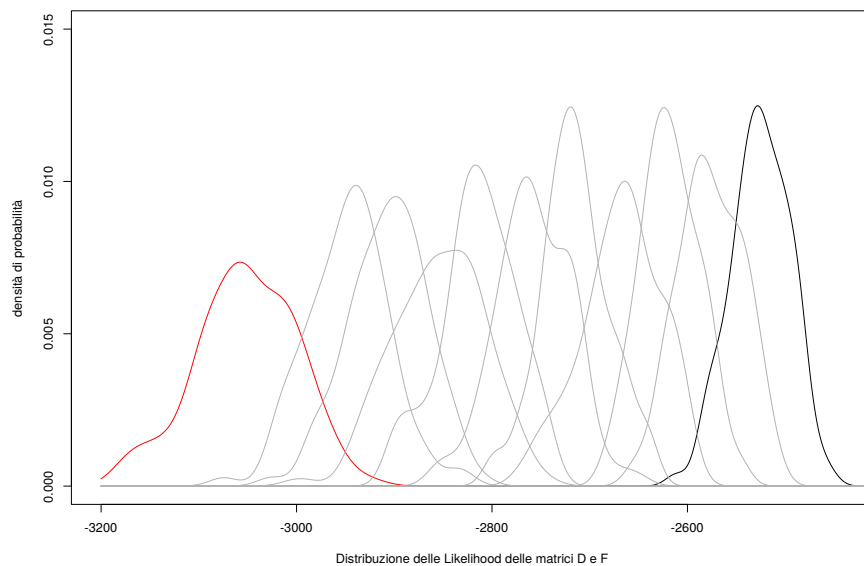


Figura 3.6: Distribuzioni delle $\log \mathcal{L}_g$ di matrici generate da modello, in nero, completamente aberrata, in rosso, e con percentuali di fakers, in grigio

Come ultimo esempio viene mostrata la Figura 3.7 che confronta le $\log \mathcal{L}_g$ della matrice non corrotta \mathbf{D} , in nero, nella matrice completamente manipolata \mathbf{F} , in rosso, e della matrice completamente casuale \mathbf{R} , in verde. Tenendo come riferimento le matrici generate da modello, si può osservare una grandissima differenza (circa 9 volte maggiore) che le manipolazioni casuali generano sulla $\log \mathcal{L}_g$, rispetto alla perturbazione generata da manipolazioni faking.

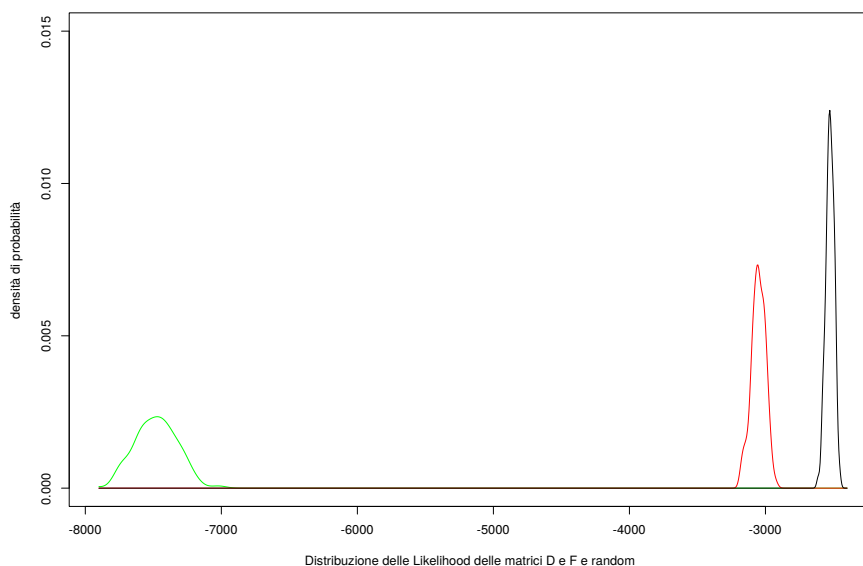


Figura 3.7: Distribuzioni delle $\log \mathcal{L}_g$ di matrici generate da modello, in nero, completamente aberrate, in rosso, e casuali, in verde

3.5 Conclusioni

I risultati della simulazione MC non sono soddisfacenti, nel senso che non riteniamo sufficientemente affidabili gli indici PFS nell'identificare i soggetti fakers, invece sembrano esserlo per individuare i pattern casuali. Nell'ultima parte del capitolo abbiamo cercato di dare un significato alla bassa potenza dei PFS nel caso di pattern fake, nonostante gli evidenti limiti tecnici della piccola simulazione, l'utilizzo di un caso critico conferma i dubbi sull'affidabilità emersi nella sezione precedente. Unendo i risultati della simulazione MC e l'analisi della likelihood, abbiamo il forte sospetto che i fakers vengano individuati solo quando producono pattern incoerenti al loro interno, quindi pattern improbabili e di conseguenza identificabili dai PFS. Questo si verifica in particolar modo quando la metà delle risposte viene manipolata, superato questo limite, più il soggetto mente attraverso gli item, più viene creato un pattern coerente ma con un punteggio superiore a quello onesto e con relazioni tra le variabili alterate.

I PFS sono stati ideati per identificare pattern non coerenti con il modello oppure non coerenti al loro interno, in entrambi i casi però si può riformulare il problema come infrequenza di un pattern rispetto alla norma. Ma l'aberrazione di faking non porta sempre a pattern infrequenti, anzi a volte potrebbe migliorare la coerenza, dipende dal pattern di partenza. Inoltre, visto che anche modelli di faking complessi e probabilistici passano molto facilmente la scrematura fatta dai PFS, l'utilizzo di modelli

molto semplici di faking (ad esempio aggiungere una valore costante a tutti gli item) dovrebbero passare ancora di più inosservati perché interferiscono meno sulla coerenza della struttura delle risposte e portano al mentitore i vantaggi desiderati.

Capitolo 4

Analisi formale dei modelli di mixture SGR

4.1 L'SGR a livello di popolazione, mixture SGR

In questo capitolo approfondiremo le potenzialità del M-SGR (Mixture Sample Generation by Replacement) da un punto di vista della sua rappresentazione analitica. Questo capitolo si discosta dai precedenti capitoli in quanto non contiene un'applicazione pratica, ma costituisce esclusivamente un contributo di analisi formale/strutturale per lo sviluppo e risoluzione delle equazioni alla base del modello M-SGR. L'analisi sarà limitata ai soli casi elementari monovariati (1 item) e bivariati (2 item) associati a variabili discrete con supporto finito in quanto per ragioni di complessità computazionale il caso multivariato è al momento difficilmente trattabile. Infatti vi è un grosso limite computazionale che fa spesso preferire nella pratica l'utilizzo dell'approccio simulativo SGR alla sua controparte analitica. Nello specifico, la difficoltà principale dell'approccio analitico dell'SGR è relativa alla necessità di derivare le distribuzioni marginali del modello di risposta che, nelle strutture di dati multivariati, può facilmente presentare il cosiddetto problema del *curse of dimensionality* (ovvero l'esplosione combinatoria dello spazio di dati da considerare per ottenere i valori delle distribuzioni marginali). In pratica, l'approccio analitico necessita che siano calcolate tutte le possibili combinazioni di risposta sia nel caso di risposta onesta che risposta manipolata. Per esempio, immaginiamo una scala verosimile con 10 item, dove ogni item ha $v = 5$ livelli di risposta. In questo caso vi sono un totale di 9,765,625 possibili combinazioni di risposte oneste. Nonostante il problema tremendo del *curse of dimensionality* che affligge la rappresentazione analitica dell'SGR, è in ogni caso sensato ed interessante individuare quali siano le equazioni fondamentali che collegano i termini in gioco nella rappresentazione del M-SGR sebbene tali formulazioni analitiche siano limitate ai casi

elementari monovariato e bivariato.

L'assunzione che sta dietro all'M-SGR, così come descritta a seguire, è che ci sia una proporzione di soggetti ($\alpha \in [0, 1]$) i quali manipolano le risposte e sono omogenei nell'utilizzo del modello di sostituzione (alias il modello di faking, la matrice di sostituzione). Questa procedura potrebbe aprire nuove opportunità nell'utilizzo della metodologia SGR poiché permette di calcolare direttamente la distribuzione di probabilità marginale associata al modello misto (costituito da una parte di individui onesti e da una parte di fakers). Se gli assunti sono corretti, tale distribuzione di probabilità permetterebbe di calcolare molte statistiche della popolazione mista invece di approssimarle tramite apposite simulazioni di tipo MC utilizzando il paradigma standard SGR. Inoltre, una volta derivata la distribuzione marginale del modello misto si potrebbe stimare direttamente le statistiche campionarie della popolazione mista generando campioni direttamente da essa, senza dover passare attraverso la procedura a due step (simulazione dei dati onesti e successiva loro perturbazione) dell'approccio SGR standard. In particolare, ci soffermeremo sulle interessanti informazioni che potrebbero essere derivate matematicamente dall'utilizzo delle manipolazioni algebriche dei termini che costituiscono il modello misto di risposta.

4.1.1 Caso univariato

Introduciamo adesso la notazione che verrà utilizzata in questa sezione. Sia $v \in \mathbb{N}$ il numero di livelli di risposta dell'item discreto ordinale. Indichiamo con $\mathbf{p}_D \in \mathcal{P}^{1 \times v}$ il vettore riga che rappresenta le v probabilità marginali di risposta secondo il modello onesto i cui parametri sono denominati come $\boldsymbol{\theta}_D$ e dove $\mathcal{P}^{1 \times v}$ sta ad indicare lo spazio di tutte le distribuzioni di massa univariate con v categorie. Le celle del vettore $\mathbf{p}_D = [p_1^d, p_2^d, \dots, p_v^d]$ rappresentano le probabilità marginali di osservare una certa risposta onesta $p(x|\boldsymbol{\theta}_D)$ con $x = 1, \dots, v$ nel modello onesto. In modo analogo indichiamo con $\mathbf{p}_F \in \mathcal{P}^{1 \times v}$ il vettore riga che rappresenta le v probabilità marginali di risposta secondo il modello puro di faking i cui parametri sono indicati dal vettore $\boldsymbol{\theta}_F$. Le celle di $\mathbf{p}_F = [p_1^f, p_2^f, \dots, p_v^f]$ rappresentano le probabilità marginali di osservare una certa risposta fake $p(x|\boldsymbol{\theta}_F)$ con $x = 1, \dots, v$. Infine $\mathbf{p}_X \in \mathcal{P}^{1 \times v}$ sta ad indicare l'array di probabilità generato da una somma pesata tra i vettori \mathbf{p}_D e \mathbf{p}_F . Ogni cella di \mathbf{p}_X rappresenta la probabilità marginale di una risposta nella popolazione mista, $p(x|\boldsymbol{\theta}_D, \boldsymbol{\theta}_F, \alpha)$ dove lo ricordiamo α denota la proporzione di fakers nella popolazione mista.

La relazione tra \mathbf{p}_X , \mathbf{p}_F e \mathbf{p}_D può essere formalizzata dalla seguente formula:

$$\mathbf{p}_X = \mathbf{p}_D(1 - \alpha) + \mathbf{p}_F \quad (4.1)$$

dove:

$$\mathbf{p}_F = \mathbf{p}_D \mathbf{Z} \quad (4.2)$$

dove ricordiamo che \mathbf{Z} è chiamata nella terminologia SGR la *replacement matrix*, matrice di sostituzione, più in generale è chiamata matrice stocastica, o di transizione. Essa contiene valori di probabilità condizionali le cui righe sommano a 1. Più precisamente \mathbf{Z} rappresenta le probabilità condizionate di $p(x|x', \boldsymbol{\theta}_F)$ dove x è il valore sostituito e x' è il vecchio valore. \mathbf{Z} contiene tutte le informazioni riguardo al modello di manipolazione applicato all'item.

Dalla formula appena rappresentata si può direttamente derivare:

$$\mathbf{p}_D = \mathbf{p}_F \mathbf{Z}^{-1}; \quad (4.3)$$

la quale risponde alla domanda: se avessimo osservato un campione completamente manipolato (cioè $\alpha = 1$ e di conseguenza $\mathbf{p}_X = \mathbf{p}_F$), quale sarebbe l'esito di \mathbf{p}_D ?

Sostituendo nella formula 4.1, \mathbf{p}_F con $\mathbf{p}_D \mathbf{Z}$ otteniamo:

$$\mathbf{p}_X = \mathbf{p}_D(1 - \alpha) + \mathbf{p}_D \mathbf{Z} \alpha. \quad (4.4)$$

Simmetricamente dato \mathbf{s}_X e considerata un'ipotesi di faking (la quale è composta da due parti: proporzione di fakers α e dal modello di faking \mathbf{Z}), si può calcolare \mathbf{p}_D :

$$\mathbf{p}_D = \mathbf{p}_X [(1 - \alpha) \mathbf{I}_v + \mathbf{Z} \alpha]^{-1}. \quad (4.5)$$

Invece, sostituendo nella formula 1, \mathbf{p}_D con $\mathbf{p}_F \mathbf{Z}^{-1}$:

$$\mathbf{p}_F = \mathbf{p}_X [\mathbf{Z}^{-1}(1 - \alpha) + \mathbf{I}_v \alpha]^{-1}; \quad (4.6)$$

la quale ci dà la possibilità di calcolare le probabilità di \mathbf{p}_F osservando \mathbf{p}_X e avendo un'ipotesi sul modello di faking. Si noti che nelle equazioni sopra riportate \mathbf{I}_v sta ad indicare la matrice identità di ordine v .

Inoltre, osservando \mathbf{p}_X e conoscendo almeno uno tra \mathbf{p}_D e \mathbf{p}_F si può stimare la percentuale fakers nella popolazione:

$$\alpha = (\mathbf{p}_X - \mathbf{p}_D)(-\mathbf{p}_D + \mathbf{p}_F)^{-1} \quad (4.7)$$

Infine si può provare a calcolare direttamente \mathbf{Z} , ma purtroppo la matrice che si ottiene

sottende indipendenza tra le risposte oneste e quelle manipolate:

$$\mathbf{Z} = \mathbf{p}_D^{-1} \mathbf{p}_F \quad (4.8)$$

e questo può condurre ad un risultato indeterminato dove l'esito è una matrice simile a \mathbf{Z} , funzionalmente equivalente, ma senza il significato che usualmente si attribuisce a \mathbf{Z} .

4.1.2 Caso bivariato

Nel caso bivariato indichiamo con v_1 e v_2 il numero di possibili risposte discrete (ordinali) associate all'item 1 e all'item 2. Utilizzando gli stessi assunti della sezione precedente, cioè che solo due popolazioni vadano a comporre la mistura, la popolazione onesta e la popolazione dei manipolatori le quali sono descritte rispettivamente dalle matrici di probabilità marginale congiunta $\mathbf{\Pi}_D$ ($v_1 \times v_2$) e $\mathbf{\Pi}_F$ ($v_1 \times v_2$), e che la popolazione dei manipolatori sia omogenea nel modello di manipolazione, allora si può calcolare la probabilità congiunta della popolazione mista $\mathbf{\Pi}_X$ ($v_1 \times v_2$) tra queste due popolazioni.

Più precisamente, $\mathbf{\Pi}_D \in \mathcal{P}^{v_1 \times v_2}$ rappresenta la matrice delle probabilità marginali congiunte di risposta generata secondo il modello onesto i cui parametri sono denominati $\boldsymbol{\theta}_D$. Ogni cella di $\mathbf{\Pi}_D$ rappresenta la probabilità di osservare un certo pattern onesto $p(x_1, x_2 | \boldsymbol{\theta}_D)$. In modo analogo, $\mathbf{\Pi}_F \in \mathcal{P}^{v_1 \times v_2}$ rappresenta la matrice delle probabilità di risposta marginali congiunte generata aberrando la matrice $\mathbf{\Pi}_D$, oppure l'array di probabilità di risposta marginali congiunte generate dal modello che segue i parametri $\boldsymbol{\theta}_F$. Ogni cella di $\mathbf{\Pi}_F$ rappresenta la probabilità di osservare un certo pattern manipolato $p(x_1, x_2 | \boldsymbol{\theta}_F)$. Infine, $\mathbf{\Pi}_X \in \mathcal{P}^{v_1 \times v_2}$ sta ad indicare la matrice di probabilità marginali congiunte generata da una somma pesata tra le matrici $\mathbf{\Pi}_D$ e $\mathbf{\Pi}_F$. Ogni cella di $\mathbf{\Pi}_X$ rappresenta la probabilità marginale di osservare un pattern nella popolazione mista, $p(x_1, x_2 | \boldsymbol{\theta}_D, \boldsymbol{\theta}_F, \alpha)$.

La relazione tra $\mathbf{\Pi}_X$, $\mathbf{\Pi}_D$ e $\mathbf{\Pi}_F$ può essere formalizzata dalla seguente formula:

$$\mathbf{\Pi}_X = \mathbf{\Pi}_D(1 - \alpha) + \mathbf{\Pi}_F\alpha \quad (4.9)$$

dove:

$$\mathbf{\Pi}_F = (\mathbf{\Pi}_D^T \mathbf{Z}_1)^T \mathbf{Z}_2 \quad (4.10)$$

\mathbf{Z}_j è la *replacement matrix* applicata sull'item j (con $j = 1, 2$).

$$\mathbf{\Pi}_F = ((\mathbf{\Pi}_D)^T \mathbf{Z}_1)^T \times \mathbf{Z}_2 \quad (4.11)$$

Già nella situazione con due soli item si complica di molto il calcolo delle funzioni inverse. Per semplificare questo procedimento inseriamo degli elementi di algebra vettoriale: la funzione di vettorializzazione di una matrice $\vec{\cdot}$ e il prodotto di Kronecker \otimes . La vettorializzazione trasforma una matrice in un vettore inserendo i dati colonna per colonna. Il prodotto di Kronecker è un particolare tipo di operazione tra due matrici: date due matrici \mathbf{A} di ordine $m \times n$ e \mathbf{B} di ordine $p \times q$, $\mathbf{A} \otimes \mathbf{B}$ moltiplica ogni elemento di \mathbf{A} per tutta la matrice \mathbf{B} ottenendo una matrice $mp \times nq$.

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}$$

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & \cdots & a_{11}b_{1q} & \cdots & \cdots & a_{1n}b_{11} & a_{1n}b_{12} & \cdots & a_{1n}b_{1q} \\ a_{11}b_{21} & a_{11}b_{22} & \cdots & a_{11}b_{2q} & \cdots & \cdots & a_{1n}b_{21} & a_{1n}b_{22} & \cdots & a_{1n}b_{2q} \\ \vdots & \vdots & \ddots & \vdots & & & \vdots & \vdots & \ddots & \vdots \\ a_{11}b_{p1} & a_{11}b_{p2} & \cdots & a_{11}b_{pq} & \cdots & \cdots & a_{1n}b_{p1} & a_{1n}b_{p2} & \cdots & a_{1n}b_{pq} \\ \vdots & \vdots & & \vdots & \ddots & & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & & \ddots & \vdots & \vdots & & \vdots \\ a_{m1}b_{11} & a_{m1}b_{12} & \cdots & a_{m1}b_{1q} & \cdots & \cdots & a_{mn}b_{11} & a_{mn}b_{12} & \cdots & a_{mn}b_{1q} \\ a_{m1}b_{21} & a_{m1}b_{22} & \cdots & a_{m1}b_{2q} & \cdots & \cdots & a_{mn}b_{21} & a_{mn}b_{22} & \cdots & a_{mn}b_{2q} \\ \vdots & \vdots & \ddots & \vdots & & & \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{p1} & a_{m1}b_{p2} & \cdots & a_{m1}b_{pq} & \cdots & \cdots & a_{mn}b_{p1} & a_{mn}b_{p2} & \cdots & a_{mn}b_{pq} \end{bmatrix}.$$

Quindi, partendo da:

$$\vec{\Pi}_X = \vec{\Pi}_D(1 - \alpha) + \vec{\Pi}_F\alpha; \quad (4.12)$$

dove

$$\vec{\Pi}_F = \mathbf{Z}_2^T \otimes \mathbf{Z}_1^T \vec{\Pi}_D; \quad (4.13)$$

si può derivare che:

$$\vec{\Pi}_D = (\mathbf{Z}_2^T \otimes \mathbf{Z}_1^T)^{-1} \vec{\Pi}_F; \quad (4.14)$$

la quale risponde alla domanda: se avessimo osservato un campione completamente manipolato (cioè $\alpha = 1$ e di conseguenza $\mathbf{\Pi}_X = \mathbf{\Pi}_F$), quale sarebbe $\mathbf{\Pi}_D$?

Sostituendo, nella formula 4.14, $\vec{\mathbf{\Pi}}_F$ con $\mathbf{Z}_2^T \otimes \mathbf{Z}_1^T \vec{\mathbf{\Pi}}_D$ otteniamo:

$$\vec{\mathbf{\Pi}}_X = \vec{\mathbf{\Pi}}_D(1 - \alpha) + \mathbf{Z}_2^T \otimes \mathbf{Z}_1^T \vec{\mathbf{\Pi}}_D \alpha. \quad (4.15)$$

Se si osserva empiricamente $\mathbf{\Pi}_X$ e si ha un'ipotesi di faking (la quale è composta da: percentuale di fakers e dai modelli di faking \mathbf{Z}_1 e \mathbf{Z}_2), si può calcolare $\vec{\mathbf{\Pi}}_D$:

$$\vec{\mathbf{\Pi}}_D = [(1 - \alpha \mathbf{I}_{vv}) + \mathbf{Z}_2^T \otimes \mathbf{Z}_1^T \alpha]^{-1} \vec{\mathbf{\Pi}}_X. \quad (4.16)$$

Invece sostituendo nella formula 1 $\vec{\mathbf{\Pi}}_D$ con $(\mathbf{Z}_2^T \otimes \mathbf{Z}_1^T)^{-1} \vec{\mathbf{\Pi}}_F$:

$$\vec{\mathbf{\Pi}}_F = [(\mathbf{Z}_2^T \otimes \mathbf{Z}_1^T)^{-1}(1 - \alpha) + \mathbf{I}_{vv} \alpha]^{-1} \vec{\mathbf{\Pi}}_X; \quad (4.17)$$

la quale ci dà la possibilità di calcolare le probabilità di $\mathbf{\Pi}_F$ osservando $\mathbf{\Pi}_X$ e avendo un'ipotesi sul modello di faking.

Osservando $\mathbf{\Pi}_X$ e conoscendo almeno uno tra $\mathbf{\Pi}_D$ e $\mathbf{\Pi}_F$ si può calcolare la percentuale fakers nella popolazione:

$$\alpha = (\mathbf{\Pi}_x - \mathbf{\Pi}_D)(-\mathbf{\Pi}_D + \mathbf{\Pi}_F)^{-1} \quad (4.18)$$

Infine, non ci siamo addentrati a calcolare \mathbf{Z}_1 e \mathbf{Z}_2 date le problematiche non risolte già con un singolo item.

Di seguito sono schematizzati le principali derivazioni delle formule presentate in questa sezione.

4.2 Derivazioni Analitiche

4.2.1 1 item

Definite queste relazioni:

$$\begin{cases} \mathbf{p}_X = \mathbf{p}_D(1 - \alpha) + \mathbf{p}_F \alpha \\ \mathbf{p}_F = \mathbf{p}_D \mathbf{Z} \\ \mathbf{p}_D = \mathbf{p}_F \mathbf{Z}^{-1} \end{cases} \quad (4.19)$$

Calcolo di \mathbf{p}_D

$$\mathbf{p}_X = \mathbf{p}_D(1 - \alpha) + \mathbf{p}_D\mathbf{Z}\alpha \quad (4.20)$$

$$\mathbf{p}_X = \mathbf{p}_D[(1 - \alpha\mathbf{I}_q) + \mathbf{Z}\alpha] \quad (4.21)$$

$$\mathbf{p}_D = \mathbf{p}_X[(1 - \alpha\mathbf{I}_q) + \mathbf{Z}\alpha]^{-1} \quad (4.22)$$

Calcolo di $\boldsymbol{\pi}_F$

$$\mathbf{p}_X = \mathbf{p}_F\mathbf{Z}^{-1}(1 - \alpha) + \mathbf{p}_F\alpha \quad (4.23)$$

$$\mathbf{p}_X = \mathbf{p}_F[\mathbf{Z}^{-1}(1 - \alpha) + \mathbf{I}_v\alpha] \quad (4.24)$$

$$\mathbf{p}_F = \mathbf{p}_X[\mathbf{Z}^{-1}(1 - \alpha) + \mathbf{I}_v\alpha]^{-1} \quad (4.25)$$

Calcolo di α

$$\mathbf{p}_X = \mathbf{p}_D - \mathbf{p}_D\alpha + \mathbf{p}_F\alpha \quad (4.26)$$

$$\mathbf{p}_X = \mathbf{p}_D + \alpha(-\mathbf{p}_D + \mathbf{p}_F) \quad (4.27)$$

$$\alpha = (\mathbf{p}_X - \mathbf{p}_D)(-\mathbf{p}_D + \mathbf{p}_F)^{-1} \quad (4.28)$$

Calcolo di \mathbf{Z}

$$\mathbf{p}_X = \mathbf{p}_D(1 - \alpha) + \mathbf{p}_D\mathbf{Z}\alpha \quad (4.29)$$

$$(\mathbf{p}_D\mathbf{Z})^{-1} = \left[\frac{\mathbf{p}_X - \mathbf{p}_D(1 - \alpha)}{\alpha} \right]^{-1} \quad (4.30)$$

$$\mathbf{Z}^{-1}\mathbf{p}_D^{-1} = \mathbf{p}_F^{-1} \quad (4.31)$$

$$\mathbf{Z}^{-1} = \mathbf{p}_F^{-1}\mathbf{p}_D \quad (4.32)$$

$$\mathbf{Z} = (\mathbf{p}_F^{-1}\mathbf{p}_D)^{-1} \quad (4.33)$$

$$\mathbf{Z} = \mathbf{p}_D^{-1}\mathbf{p}_F \quad (4.34)$$

4.3 2 item

$$\left\{ \begin{array}{l} \boldsymbol{\Pi}_X = \boldsymbol{\Pi}_D(1 - \alpha) + \boldsymbol{\Pi}_F\alpha \\ \boldsymbol{\Pi}_F = (\boldsymbol{\Pi}_D^T\mathbf{Z}_1)^T\mathbf{Z}_2 = \mathbf{Z}_1^T\boldsymbol{\Pi}_D\mathbf{Z}_2 \\ \boldsymbol{\Pi}_D = (\boldsymbol{\Pi}_F^T\mathbf{Z}_1^{-1})^T\mathbf{Z}_2^{-1} = (\mathbf{Z}_1^{-1})^T\boldsymbol{\Pi}_F\mathbf{Z}_2^{-1} \\ \vec{\boldsymbol{\Pi}}_F = \mathbf{Z}_2^T \otimes \mathbf{Z}_1^T \vec{\boldsymbol{\Pi}}_D \\ \vec{\boldsymbol{\Pi}}_D = (\mathbf{Z}_2^T \otimes \mathbf{Z}_1^T)^{-1} \vec{\boldsymbol{\Pi}}_F \end{array} \right. \quad (4.35)$$

Calcolo di $\mathbf{\Pi}_D$

$$\vec{\mathbf{\Pi}}_X = \vec{\mathbf{\Pi}}_D(1 - \alpha) + \mathbf{Z}_2^T \otimes \mathbf{Z}_1^T \vec{\mathbf{\Pi}}_D \alpha \quad (4.36)$$

$$\vec{\mathbf{\Pi}}_X = [(\alpha \mathbf{I}_{qq}) + \mathbf{Z}_2^T \otimes \mathbf{Z}_1^T \alpha] \vec{\mathbf{\Pi}}_D \quad (4.37)$$

$$\vec{\mathbf{\Pi}}_D = [(1 - \alpha \mathbf{I}_{qq}) + \mathbf{Z}_2^T \otimes \mathbf{Z}_1^T \alpha]^{-1} \vec{\mathbf{\Pi}}_X \quad (4.38)$$

Calcolo di $\mathbf{\Pi}_F$

$$\vec{\mathbf{\Pi}}_X = (\mathbf{Z}_2^T \otimes \mathbf{Z}_1^T)^{-1} \vec{\mathbf{\Pi}}_F(1 - \alpha) + \vec{\mathbf{\Pi}}_F \alpha \quad (4.39)$$

$$\vec{\mathbf{\Pi}}_X = [(\mathbf{Z}_2^T \otimes \mathbf{Z}_1^T)^{-1}(1 - \alpha) + \mathbf{I}_{qq} \alpha] \vec{\mathbf{\Pi}}_F \quad (4.40)$$

$$\vec{\mathbf{\Pi}}_F = [(\mathbf{Z}_2^T \otimes \mathbf{Z}_1^T)^{-1}(1 - \alpha) + \mathbf{I}_{qq} \alpha]^{-1} \vec{\mathbf{\Pi}}_X \quad (4.41)$$

Calcolo di α

$$\vec{\mathbf{\Pi}}_X = \vec{\mathbf{\Pi}}_D - \vec{\mathbf{\Pi}}_D \alpha + \vec{\mathbf{\Pi}}_F \alpha \quad (4.42)$$

$$\vec{\mathbf{\Pi}}_X = \vec{\mathbf{\Pi}}_D + \alpha(-\vec{\mathbf{\Pi}}_D + \vec{\mathbf{\Pi}}_F) \quad (4.43)$$

$$\alpha = (-\vec{\mathbf{\Pi}}_D + \vec{\mathbf{\Pi}}_F)^{-1}(\vec{\mathbf{\Pi}}_X - \vec{\mathbf{\Pi}}_D) \quad (4.44)$$

Capitolo 5

Discussione Generale

In fase di revisione della tesi ci è stato fatto notare che non è riportata un'applicazione empirica. In verità, noi abbiamo provato a farla qualche tempo fa, ha aperto complicazioni e sollevato problemi che al momento non sappiamo ancora come risolvere completamente, abbiamo preferito metterla da parte per concludere il resto del lavoro, forse oggi meriterebbe un approfondimento e un accurato lavoro indipendente. Cerchiamo intanto di ricostruire cronologicamente il processo di ricerca che ha portato alla scrittura di questa tesi.

Il lavoro del capitolo 3 è stato il primo a iniziare e a essere terminato esso ci ha permesso di risollevere il problema dell'identificazione dei fakers e di definire quale riteniamo essere la motivazione principale della poca efficienza degli indici di fit individuale (PFS), cioè una non discriminazione tra i fakers e gli onesti sulla base della likelihood. Le somiglianze con altri tipi di manipolazioni contenute nella parte di review contenuta nello stesso capitolo (par. 3.2) ci hanno dato conferma, almeno a nostro avviso, che l'SGR si inserisce di diritto all'interno delle manipolazioni dei dati. Oltretutto non si limita al faking, consideriamo il faking solo uno dei possibili impieghi, basta utilizzare un adeguato modello manipolativo a seconda della specifica situazione.

Se da un lato la simulazione per individuare i fakers è stata piuttosto robusta e approfondita (par. 3.3), dall'altro la parte conclusiva riguardo lo studio della likelihood (par. 3.4), purtroppo, è stata trattata più velocemente solo attraverso esempi e la generalizzazione a tutti i questionari solo dedotta. Probabilmente sarebbe interessante approfondire quella sezione per farla diventare un robusto lavoro indipendente. Nella stesura della simulazione era già emerso il concetto della mixture, cioè che il comportamento del mentitore sia coerente all'interno del pattern di risposte e che il campione si suddivida in sinceri e fakers, ma non ulteriormente formalizzato. La situazione migliore per il mentitore è quando mente su quasi la totalità degli item, cercando di creare pattern menzogneri, ma coerenti e questo comportamento gli crea un vantaggio nel non

essere identificato (figure da 3.1 a 3.4) rispetto al mentire solo a metà degli item.

La seconda fase del percorso che sta dietro alla tesi è una piccola raccolta dati che abbiamo preferito non trattare nel lavoro finale. Purtroppo non avevamo trovato risultati sufficientemente completi ed interessanti per la comunità scientifica, ma ci ha dato spunti di ragionamento che poi hanno portato alla descrizione della Mixture-SGR. Quindi, accenniamo brevemente l'esperimento perché contiene una somministrazione insolita, ma innovativa, coerente con il modello SGR, e riportiamo alcuni appunti interessanti che hanno fatto da motore per la successiva fase (cap. 2) la quale racchiude sinteticamente i significati più importanti che siamo riusciti ad identificare e che volevamo trasmettere occupandoci di faking in questo frangente. Abbiamo chiesto a circa 100 soggetti di rispondere al questionario SCL-90-R (una scala di sintomi psicosomatici, Prunas, Sarno, Preti, Madeddu, & Perugini, 2012), la prima volta onestamente e la seconda immaginandosi in una situazione in cui avrebbero ottenuto dei benefici monetari se avessero mostrato sufficienti sintomi in modo da essere nel primo 10° percentile dei partecipanti, ma con la minaccia di eventuali verifiche mediche.

La particolarità di questa somministrazione è il fatto che abbiamo chiesto di rispondere alle due somministrazioni una di seguito all'altra e sullo stesso foglio di risposta, in modo che la risposta onesta fosse saliente e visibile al soggetto, questo per rendere la risposta onesta rilevante ed essere coerente con il modello SGR. Altri tipi di somministrazione introducono bias, ad esempio di memoria (il soggetto non ricorda la risposta vera che aveva dato la settimana prima) o situazionali (effettivamente la sintomatologia è cambiata nel soggetto nel corso del tempo), non facilmente controllabili dallo sperimentatore. Proponiamo questo tipo di doppia somministrazione nella speranza possa essere utile a futuri studi simili. È emerso che tutti avevano intuito la direzione in cui mentire, anche aiutati dal fatto che gli item siano molto espliciti sul significato che vanno ad indagare. Un altro aspetto interessante sta nel fatto che i soggetti si sono divisi principalmente in due gruppi: chi aumentava di tanto e chi aumentava di poco il punteggio test, idea poi formalizzata con la formula 2.16, la quale descrive più modelli di faking all'interno della stessa popolazione.

Rimanendo però con l'idea della non-indipendenza tra item, con l'obiettivo di innovare l'utilizzo della struttura SGR fino a quel momento utilizzata, ci siamo resi conto che un questionario di molti item è difficile da analizzare perché le combinazioni di risposte sono esponenziali al numero di item. Quindi abbiamo scelto di ricominciare il percorso analitico da due item per non perdere il nostro assunto di interdipendenza nelle risposte. La decisione di questo "passo indietro" si è mostrata vincente perché ci avrebbe successivamente permesso l'esplorazione matematica delle distribuzioni di probabilità nel caso bivariato. Purtroppo resta ancora difficile applicare le formule

(cap. 4) a casi reali per la difficoltà a ipotizzare i veri modelli di manipolazione e la quantità di soggetti faker, ad esempio, per cercare di ricostruire i dati originali (formula 4.16). Nonostante i limiti empirici a livello campionario, lo studio matematico ha permesso lo studio a livello di popolazione e mostrare che gli effetti del faking possono diventare facilmente invasivi nell'esistenza di una ricerca gonfiando o diminuendo correlazioni o, addirittura, invertendone il segno. In particolare, i nostri risultati si possono sovrapporre, almeno parzialmente, a quelli di Loken e Gelman (2017). Nell'articolo gli autori parlano di errore casuale di misura, il quale può modificare la dimensione dell'effetto principale, anche accentuandolo nei piccoli campioni. Siamo arrivati a conclusioni simili quasi contemporaneamente; anzi il nostro tipo di "errore" di misura è più evoluto e rappresenta un comportamento complesso, strategico e non indipendente (attraverso gli item) dei soggetti. In particolare gli autori ci servono su un piatto d'argento la giustificazione di tutte le ore di lavoro che abbiamo dedicato a questo progetto: "*The situation become more complicated in problems with multiple predictors, or with nonindependent error*" (pg. 585). Il faking potrebbe essere, appunto, un esempio di errore non-indipendente. Ovviamente buoni presupposti non sono garanzia di un ottimo risultato, ma sono almeno un buon punto di partenza.

Qualcuno potrebbe criticare i forti assunti modellistici riguardo al faking dai quali dipendono i risultati, cioè i modelli *sligth*, *average* ed *extreme*. Dal nostro punto di vista, i modelli di faking sono modificabili, personalizzabili in base alle esigenze dello sperimentatore; in ogni caso, riteniamo modelli utilizzati sufficientemente credibili. Oltretutto, riteniamo che i risultati siano abbastanza robusti rispetto ai modelli di faking utilizzati, cioè anche con modelli diversi (ma che rappresentino il faking) i risultati saranno simili; le tendenze degli effetti sono simili tra *sligth*, *average* ed *extreme*, cambiano invece di intensità negli effetti. Purtroppo testare la validità di modelli è difficile, in particolare se sono così dettagliati, quanti soggetti servirebbero? Se qualcuno ha delle proposte convincenti, e risorse, saremo ben lieti di lavorarci assieme.

Cambiando argomento e parlando dell'applicabilità dell'impostazione Mixture, sono state condotte delle analisi esplorative su dati sensibili probabilmente corrotti da faking, elaborati in $\mathbf{\Pi}_X$ e si è cercato di ricostruire la matrice $\mathbf{\Pi}_D$, purtroppo questa strategia si è rivelata, almeno nei pochi casi da noi testati, inadeguata per limiti empirici: cambiando i parametri di faking in maniera sistematicamente esplorativa, si ottengono solo dati parziali e troppo aleatori per fare delle stime affidabili. Inoltre, le eventuali, ma frequenti empiricamente, celle vuote della matrice $\mathbf{\Pi}_X$ creano impossibilità di calcolo. Una diversa strategia potrebbe essere quella di aggirare le matrici delle probabilità congiunta ($\mathbf{\Pi}$) passando dalla dimensione dell'effetto Q . Cioè, ipotizzando due verosimili distribuzioni marginali oneste delle due variabili e un range

di ipotetiche correlazioni oneste, si potrebbe stimare tramite simulazione l'eventuale correlazione dovuta al faking originato in tale configurazione e aggiungerlo/sottrarlo all'effetto osservato.

Concludiamo riassumendo i possibili lavori futuri che ci piacerebbe approfondire. Uno studio intensivo e sistematico del comportamento della Likelihood per giustificare l'inefficienza degli indici di fit dei soggetti, e di conseguenza dell'impossibilità di individuare i mentitori con quel tipo di analisi, potrebbe essere una simulazione in cui si andranno a variare i parametri generativi degli item. Le formule analitiche illustrate nel capitolo 4 risentono al momento di troppi limiti empirici i quali ne precludono l'applicabilità. Interessante sarebbe cercare una soluzione tecnica per aggirare gli attuali limiti. Inoltre, l'estensione al più di 2 item risente dell'esplosione combinatoria e non è garantita un'efficienza computazionale rispetto alle simulazioni. Se si volesse procedere in tale direzione, dato che la metodologia necessita del calcolo delle probabilità di ogni combinazione risposte, al momento riteniamo che bisognerebbe passare attraverso una stima delle stesse, in modo da non doverla calcolare, e rendere il processo più veloce. Indispensabile dovrebbe essere continuare la verifica e l'innovazione dei modelli di faking attualmente utilizzati anche su grandi campioni. Infatti, come ultimo punto, suggeriamo la definizione, già avviata con report interni, anche un di altri due modelli di faking. Il primo modello sarebbe un adattamento degli attuali modelli di faking per risposte categoriali, nelle quali attualmente i modelli non hanno una diretta applicabilità. Riassumendolo velocemente, in una prima fase si vanno ad ordinare le risposte sulla base di svariati fattori salienti (ad esempio quattro risposte riguardo allo sport preferito potrebbero essere ordinate per pericolosità, costo dell'attività, diffusione dello sport), si calcolano le tradizionali matrici di sostituzione con i modelli di faking, una per ogni fattore saliente (cioè una per ogni ordinamento ritenuto interessante). A questo punto si possono aggregare le diverse matrici creandone una comune, magari pesandole in base all'importanza di ciascun fattore. Quest'ultima matrice aggregata rappresenta la probabilità di sostituzione e grazie a questa si estrarrebbero le risposte condizionate. Il secondo modello, è stato chiamato momentaneamente "faking verso un target". In questo modello di faking si ipotizza che la risposta più probabile, la moda, del modello di rimpiazzo sia una specifica risposta della scala. Ad esempio, in una scala da 1 a 5, il target potrebbe essere la 4 indipendentemente dalla risposta originale, mantenendo però la specificità nella distribuzione di rimpiazzo condizionata alla risposta vera tipica del SGR (cioè sulla base della distribuzione beta generalizzata). Questo dovrebbe poter simulare comportamenti di faking più evoluti da parte dei soggetti menzionati.

Appendice A

Calcolo parallelo

In questa sezione conclusiva approfondiamo il *calcolo parallelo*, si cambia completamente settore rispetto a tutto il materiale precedentemente esposto, siamo nell'ambito dell'informatica. Forse gli argomenti trattati potrebbero sembrare banali ad un programmatore informatico, ma siamo piuttosto sicuri che i lettori di questo documento ingegneri non dovrebbero essere. Non c'è l'ambizione di essere esaustivi riguardo il tema del calcolo parallelo e lasciamo agli esperti di ingegneria e informatica tutti i dettagli tecnici, ma ci è parso molto utile fornire le basi per la comprensione e l'utilizzo di questa risorsa per una psicomtria contemporanea talvolta basata su "forza bruta", cioè algoritmi di risoluzione molto lenti e pesanti. Il legame che in questo caso lega l'ambito tecnologico-informatico con la psicomtria è che spesso c'è la necessità di sfruttare al massimo le potenzialità di un computer per simulazioni, o calcoli in generale, che richiedono molto tempo per essere eseguiti dal computer.

Una definizione di calcolo parallelo può essere l'esecuzione simultanea del codice (diviso e specificamente adattato) su più microprocessori (i personal computer attuali di hanno solitamente 2 o più), o più core dello stesso microprocessore, allo scopo di aumentare le prestazioni di calcolo del sistema di elaborazione. I vantaggi sono sia di risparmio di tempo, siamo sull'ordine, almeno, di dimezzare il tempo di computazione; sia si possono utilizzare appieno le potenzialità di un calcolatore, altrimenti non sfruttate. Per fare un esempio, è come avere un supermercato con quattro casse e, nonostante ci siano molti clienti, utilizzarne solo una andando a formare una coda che si potrebbe evitare. La miglioria proposta in questa sezione, non riguarda gli algoritmi di calcolo, i quali dovrebbero essere già ottimizzati, ma siamo su un altro livello, stiamo facendo calcolare 2, 3 o più linee di calcolo contemporaneamente. Rimanendo nell'esempio, non si va a migliorare la cassa in sé ma se ne utilizza più di una alla volta distribuendo i clienti tra di esse. Il software R, largamente utilizzato nella comunità psicomtrica, non utilizza il calcolo in parallelo in modo predefinito. Alcuni pacchetti

lo utilizzano, ma in questi casi sono stati i creatori di tali pacchetti ad implementarlo.

Se finora abbiamo decantato i vantaggi del calcolo parallelo, ora dobbiamo anche evidenziarne i limiti. Sia per la scrittura di codice leggermente più complessa, sia per limiti teorici il calcolo parallelo non è sempre la scelta più efficiente. Purtroppo, dividere tra più processori il carico di lavoro, processo di smistamento, ha un costo di tempo, quindi l'efficienza del dividere il carico del lavoro tra più microprocessori diminuisce all'aumentare degli stessi.

La metodologia proposta come esempio riguardo il calcolo parallelo si basa sul pacchetto *parallel*, e la funzione *mclapply* in esso contenuta, parte dell'ambiente statistico R (R Core Team, 2017), e si divide in tre blocchi sequenziali:

- creazione di una lista, che chiamiamo per comodità *myList*, la quale contiene tutte le combinazioni di variabili utilizzate per ogni linea di calcolo;
- creazione di una funzione ad hoc, che chiamiamo per comodità *myFunction*, la quale contiene l'algoritmo di calcolo;
- utilizzo della funzione *mclapply*, la quale utilizza la funzione *myFunction* su ogni combinazione di variabili contenuta in *myList*

L'output di questa sequenza di operazioni sarà una lista, in cui ogni elemento conterrà ciò che è specificato come output di *myFunction*.

Un esempio minimale di codice R potrebbe essere il seguente:

```
# creazione della lista
library(parallel)
myMatrix = expand.grid(Q=c(5,7),
                        nitem=c(5,10,20,50),
                        nsogg=c(300,600),
                        percF=c(.1,.25,.5),
                        intF=c(.25,.40,.50,.75,.90),
                        modello=c("extreme","slight","random")
)
myList = list()
for(i in 1:nrow(myMatrix)){
  myList[[i]] = myMatrix[i,]
}

# creazione della funzione desiderata
 #(questa somma solo dei numeri pseudocasuali)
```

```

myFunction = function(x) {

  y = x["Q"]*sample(1:100,1) + x["nitem"]*sample(1:100,1) +
    x["nsogg"]*sample(1:100,1) + x["percF"]*sample(1:100,1)+
    x["intF"]*sample(1:100,1)
  return(y)
}

#check del numero di microprocessori
detectCores(logical=TRUE)
detectCores(logical=FALSE)
# utilizzo di mclapply
system.time(mclapply(myList, myFunction, mc.cores=1))
system.time(mclapply(myList, myFunction, mc.cores=2))
system.time(mclapply(myList, myFunction, mc.cores=3))

```

Questo codice genera una matrice, `myMatrix`, che poi viene trasformata in lista, `myList`. La funzione `myFunction` utilizza le variabili all'interno della lista per creare un numero casuale, non utilizza invece la variabile categoriale "modello". Poi viene eseguito un controllo del numero dei microprocessori ed infine confrontato il tempo di esecuzione delle operazioni utilizzando 3, 2 oppure 1 microprocessori.

Con la funzione `detectCores()`, sempre contenuta in `parallel`, si possono visualizzare il numero di microprocessori all'interno del computer. Ci sono due tipi di microprocessori, quelli fisici (visualizzabili settando il parametro `logical=TRUE`) e quelli virtuali (detti anche core, visualizzabili settando il parametro `logical=FALSE`). Due microprocessori fisici sono due strutture separate, mentre due, o più, virtuali potrebbero essere contenuti all'interno dello stesso microprocessore fisico. Il dividere un microprocessore fisico in più virtuali è una strategia ingegneristica per migliorare leggermente le prestazioni; ma numero di quelli fisici influisce sulle prestazioni in modo di gran lunga superiore rispetto al numero di quelli virtuali. Sconsigliamo di utilizzare il numero massimo di microprocessori durante il processo per non sovraccaricare il computer e bloccare ogni altra operazione. Grazie alla funzione `system.time()` si possono calcolare i diversi tempi di esecuzione dell'operazione, in questo caso variando il numero di microprocessori utilizzati.

Concludendo, la computazione in parallelo è un utile strumento a disposizione dello psicometrista nel momento in cui si confronta con la programmazione. Ovviamente perde di senso se si eseguono operazioni di qualche minuto, ma spesso il tempo computazionale è di qualche giorno, se non settimane, e viene ridotto notevolmente.

Appendice B

Codici di programmazione R

∞ B.1 Capitolo 2

Per una possibile replicazione delle simulazioni alleghiamo i principali codici utilizzati per la simulazione dei dati. Inoltre, speriamo possano essere utili come base per altre simulazioni. In particolare, il seguente codice genera i dati utilizzati nel capitolo 2.

```
1
2 # librerie
3
4 library(sgr)
5 library(lavaan)
6 require(optimx)
7
8 # funzioni
9
10 q.cohen=function(corF, corD)          # Calcola della Q, se (corF > corD) allora q > 0
11 {
12   zF=1/2*log((1+corF)/(1-corF))
13   zD=1/2*log((1+corD)/(1-corD))
14   q=zF-zD
15   return(q)
16 }
17
18 rm.right=function(q,model)
19 {
20   RM=replacement.matrix(Q=q,p=c(1,0),fake.model=model)
```

```

21  return(RM)
22 }
23
24 rm.left=function(q,model)
25 {
26   RM=replacement.matrix(Q=q,p=c(0,1),fake.model=model)
27   return(RM)
28 }
29
30 dati.gen=function(corT,n,q)           # generazione dei dati seguendo il modello D
31 {
32   R=matrix(c(1,corT,corT,1),2,2)      # corT e' la correlazione teorica
33   dati=data.frame(rdatagen(n,R=R,Q=q)$data)
34   return(dati)
35 }
36
37 correlations=function(dati)
38 {
39   pear=try(lavCor(dati)[3],silent=TRUE) # calcolo dei tre tipi di correlazione, 0 sta per prima dell'aberrazione
40   spea=try(cor(dati,method="spearman")[2,1],silent=TRUE)
41   poly=try(lavCor(dati,ordered=c("X1","X2"))[3],silent=TRUE)
42   COR=c(pear,spea,poly)
43   return(COR)
44 }
45
46 # SIMULAZIONE
47
48 main="results"           # radice del nome dei file per salvare i dati
49
50 Kj=1                     # probabilita' di un item di essere manipolabile, fissata ad 1 in questa simulazione
51
52 results=expand.grid(seq(.0,.9,.1), # matrice che conterra' sia i dati input per l'analisi che gli output
53                      c(-.9,-.6,-.3,0,.3,.6,.9),
54                      c(2,5,7),
55                      c(20,50,100,1000),
56                      c("slight","extreme"),
57                      c("same","opposite"),
58                      1:5)
59 names(results)=c("Ci","corT","Q","n","model","dir","rep")
60
61 results$model=as.character(results$model)
62 results$dir=as.character(results$dir)
63
64 results$errors=as.numeric(NA)
65 results$q.pear=as.numeric(NA)      # q di Cohen
66 results$q.spea=as.numeric(NA)
67 results$q.poly=as.numeric(NA)
68 results$d.pear=as.numeric(NA)     # Correlazione empirica nel campione D
69 results$d.spea=as.numeric(NA)
70 results$d.poly=as.numeric(NA)

```

```

71 results$f.pear=as.numeric(NA)           # Correlazione empirica nel campione F
72 results$f.spea=as.numeric(NA)
73 results$f.poly=as.numeric(NA)
74
75 options(warn=-1)
76 for (II in 0:399)
77 { system.time(                          # Una funzione per misurare il tempo di esecuzione, puo' essere omessa
78   for (x in 1:nrow(results))
79   {
80     dir=results$dir[x]                   # Dal dataframe results in cui ci sono tutte e combinazioni di fattori sperimentali
81     model=results$model[x]               # Vengono richiamate una ad una tutte le combinazioni delle variabili
82     n=results$n[x]
83     q=results$Q[x]
84     corT=results$corT[x]
85     Ci=results$Ci[x]
86     errors=-1
87     cor0=NA
88     cor1=NA
89     diff.cor=NA
90     if (dir=="same")                    # Prima divisione il codice e' ripetuto eccetto che per le matrici di sostituzione (RM)
91     {
92                                             # Questo e' stato fatto perche' se la direzione del faking e' la stessa basta una RM,
93                                             # invece, se la direzione del faking e' opposta, servono due RM
94                                             # la funzione SGR di rimpiazzo (rdatarepl) necessita di un numero positivo di fakers
95     {
96       RM=rm.right(q,model)              # calcolo della matrice la matrice di sostituzione
97     }
98     repeat                                # la generazione di D viene ripetuta finche' non si riescono a calcolare tutti e 3 i tipi di correlazione
99     {
100      errors=errors+1                     # contatore per il numero di rigenerazioni dei dati
101      dati=dati.gen(corT,n,q)
102      fakers=rbinom(n,1,Ci)
103
104      if (all(var(dati)!=0))
105      {
106        cor0=correlations(dati)
107
108        if (sum(fakers)!=0)                # se la somma dei fakers e' 0, partition replacement non funziona, quindi bisogna dividere i due casi
109        {
110          datif=rbind(rdatarepl(dati[fakers==1,],RM=RM,print=FALSE)$Fx,dati[fakers==0,])
111          # rdatarepl aberra ogni cella della matrice, quindi e' stato diviso il dataset in due parti, il gruppo dei fakers e il gruppo degli onesti
112          if (all(var(datif)!=0))
113            cor1=correlations(datif)
114          } else {
115            cor1=cor0
116          }
117        }
118      } if (class(cor0)=="numeric" &      # se tutte le correlazioni sono state calcolate, allora si puo' uscire dal ciclo
119            class(cor1)=="numeric" ) break
120    }

```

```

121 } else {
122     # fine della divisione in base alla direzione di faking
123     if (Ci!=0)
124     {
125         RM=rm.right(q,model)
126         RM2=rm.left(q,model)
127     }
128     repeat
129     {
130         errors=errors+1
131         dati=dati.gen(corT,n,q)
132         fakers=rbinom(n,1,Ci) # i fakers vengono estratti
133
134         if (all(var(dati)!=0))
135         {
136             cor0=correlations(dati)
137
138             if (sum(fakers)!=0) # se la somma dei fakers e' 0, partition replacement non funziona, quindi bisogna dividere i due casi
139             {
140                 datif=dati
141                 datif[,1]=rbind(rdatarepl(as.matrix(dati[fakers==1,1],sum(fakers),1),RM=RM,print=FALSE)$Fx,as.matrix(datif[fakers==0,1],n-sum(fakers),1))
142                 # rdatarepl necessita di un data.frame o una matrice in ingresso
143                 datif[,2]=rbind(rdatarepl(as.matrix(dati[fakers==1,2],sum(fakers),1),RM=RM2,print=FALSE)$Fx,as.matrix(datif[fakers==0,2],n-sum(fakers),1))
144                 if (all(var(datif)!=0))
145                     cor1=correlations(datif) } else { cor1=correlations(dati) }
146             }
147             if (class(cor0)=="numeric" & # se tutte le correlazioni sono state calcolate, allora si puo' uscire dal ciclo
148                 class(cor1)=="numeric" ) break
149
150         }
151     }
152
153     diff.cor=q.cohen(cor1,cor0)
154
155     results[x,8:17]=c(errors,t(diff.cor), t(cor0), t(cor1))
156
157 }
158 )
159
160 options(warn=0)
161 save(results,file=paste(main,II,sep=""))
162 }

```

B.2 Capitolo 3

Il seguente codice genera i dati utilizzati nel capitolo 3.

```

1 library(catR)
2 library(ltm)
3 library(sgr)
4 library(PerFit)
5 library(pROC)
6 library(beepr)
7 library(parallel)
8 library(latticeExtra)
9 library(reshape2)
10 library(msm)
11 library(Hmisc)
12 library(gridExtra)
13 library(foreach)
14 library(doParallel)
15
16 # il file e' composto da:
17 # 1 una parte in cui vengono definite le funzioni poi utilizzate
18 # 2 generazione dei dati
19 # 3 calcolo dei PFS
20 # 4 statistiche sugli indici PFS
21 # 5 grafici
22
23 # funziona con le versioni dei pacchetti aggiornati al 17 settembre 2015
24
25 rm(list=ls())
26
27 #cartella di lavoro in cui salvare tutti i file
28 wd="/home/utente/file_simulazione"
29
30 # 1 LE FUNZIONI
31
32 #funzione per generare gli item secondo GPCM
33 genItemGPCM = function(nitem,Q)
34 {
35
36   item=data.frame(matrix(NA,nitem,Q))
37
38   for(i in 1:nitem)
39   {
40     alphaj <- rlnorm(nitem, 0, .25)
41     pars=rtnorm(Q-1,0,.1519892,-.25,.25)
42     pars=pars+rtnorm(1,0,1,-1,1)
43
44     if(Q==4)
45       pars=pars+c(-1,0,1)

```

```

46   if (Q==5)
47     pars=pars+c(-1,-.5,.5,1)
48   if (Q==7)
49     pars=pars+c(-1.5,-1,-.5,.5,1,1.5)
50
51   item[i,] <- c(alphaj[i], pars)
52 }
53 name <- "alphaj"
54 for (i in 1:(Q-1))
55   name <- c(name, paste("deltaj", i, sep = ""))
56 colnames(item) <- name
57
58 return(item)
59 }
60
61 #funzione per generare i dataset di dati
62 mySimulation = function(x) {
63
64   Q=x$Q
65   nitem=x$nitem
66   nsogg=x$nsogg
67   percF=x$percF
68   intF=x$intF
69   modello=as.character(x$modello)
70
71   item=genItemGPCM(nitem,Q)
72
73   DATI=NULL
74   theta=rnorm(nsogg)
75   for (sogg in 1:nsogg)
76     DATI=rbind(DATI,genPattern(theta[sogg], item, model = "GPCM" , D = 1.702))
77   DATI=DATI+1
78   if (modello=="random") #definisco a mano le probabilita' condizionate per il modello random
79   {
80     PM=rbind(DATI[1:(nrow(DATI)*percF),1:ncol(DATI)],matrix(0,nrow(DATI)-(nrow(DATI)*percF),ncol(DATI)))
81     if (Q==4)
82     {
83       P=matrix(c(intF,0,
84                 intF/3*2,intF/3*1,
85                 intF/3*1,intF/3*2,
86                 0,intF),ncol=2,byrow=TRUE)
87       gam =c(1,1,1,1,1,1,1,1,1)
88       del =c(1,1,1,1,1,1,1,1)
89
90     } else if (Q==5) {
91       P=matrix(c(intF,0,
92                 intF/4*3,intF/4*1,
93                 intF/4*2,intF/4*2,
94                 intF/4*1,intF/4*3,
95                 0,intF),ncol=2,byrow=TRUE)

```

```

96     gam =c(1,1,1,1,1,1,1,1,1,1)
97     del =c(1,1,1,1,1,1,1,1,1,1)
98
99   } else if (Q==7) {
100     P=matrix(c(intF,0,
101               intF/6*5,intF/6*1,
102               intF/6*4,intF/6*2,
103               intF/6*3,intF/6*3,
104               intF/6*2,intF/6*4,
105               intF/6*1,intF/6*5,
106               0,intF
107             ),ncol=2,byrow=TRUE)
107     gam =c(1,1,1,1,1,1,1,1,1,1,1,1)
108     del =c(1,1,1,1,1,1,1,1,1,1,1,1)
109   }
110
111   Pparm = list(p=P,gam=gam,del=del)
112   DATArepl=try( partition.replacement(DATI,PM,Pparm=Pparm))
113
114 } else {
115   PM=rbind(matrix(1,nrow(DATI)*percF,ncol(DATI)),matrix(0,nrow(DATI)-(nrow(DATI)*percF),ncol(DATI)))
116   DATArepl=try( partition.replacement(DATI,PM,fake.model=modello,p=matrix(c(intF,0),ncol=2)))
117 }
118
119 A=DATI==DATArepl
120 B=matrix(as.numeric(A),dim(A))
121 C=rowSums(B)+nitem
122 FH2=rep("H",nsogg)
123 FH2[which(C!=0)]= "F"
124
125 dataset=try(list(DATArepl,nitem,nsogg,percF,intF,modello,Q,C,FH2))
126 names(dataset)=c('DATIrepl','nitem','nsogg','percF','intF','modello','Q','itemAberrati',"FH2")
127
128 return(dataset)
129 }
130
131 #funzione per l'analisi dei dataset di dati
132 myAnalysis = function(x)
133 {
134   datiF=as.numeric(x$DATIrepl)
135   nsogg=as.numeric(x$nsogg)
136   nitem=as.numeric(x$nitem)
137   datiF=matrix(datiF-1,nsogg,nitem)
138   Q=x$Q
139
140   Lz=try(lzpoly(datiF,Ncat=Q,IRT.PModel="GPCM")$PFscores)
141   U3=try(U3poly(datiF,Ncat=Q)$PFscores)
142   Gp=try(Gpoly(datiF,Ncat=Q)$PFscores)
143   Gpn=try(Gnormed.poly(datiF,Ncat=Q)$PFscores)
144
145   ris=list(x$nitem,x$nsogg,x$percF,x$intF, x$modello,Q,Lz,U3,Gp,Gpn,x$FH2)

```

```

146  names(ris)= c("nitem", "nsogg", "percF", "intF", "modello", "Q", "Lz", "U3", "Gp", "Gpn", "fh")
147  return(ris)
148 }
149
150 #generazione delle combinazioni delle variabili
151 listCases=list()
152 i=0
153 for (Q in c(5,7))
154   for (nitem in c(5,10,20))
155     for (nsogg in c(300,600,900))
156       for (percF in c(.05,.10,.25))
157         for (intF in c(.25,.50,.75,.9))
158           for (modello in c("extreme", "slight", "random"))
159             {
160               i= i+1
161               listCases[[i]]= list(Q, nitem, nsogg, percF, intF, modello)
162               names(listCases[[i]])=c("Q", "nitem", "nsogg", "percF", "intF", "modello")
163             }
164
165 # 2 GENERAZIONE DEI DATASET
166 for (kk in 1:100)
167   {
168     yout=NULL
169     for (yy in 1:10)
170       {
171         timeA=Sys.time()
172         print(c(kk,yy))
173         yout = c(yout, mclapply(listCases, mySimulation, mc.cores=detectCores()))
174         timeB=Sys.time()
175         time=timeB-timeA
176         print(time)
177       }
178     save(yout, file=paste(wd, "yout", kk, sep=""))
179   }
180
181 # 3 CALCOLO DEI PFS
182 for (kk in 1:100)
183   {
184     load(file=paste(wd, "yout", kk, sep=""))
185
186     timeA=Sys.time()
187     ris = mclapply(yout, myAnalysis, mc.cores=detectCores())
188     timeB=Sys.time()
189     time=timeB-timeA
190     print(kk)
191     print(time)
192
193     #tolgo tutti i dataset dove il modello IRT non ha fittato
194     j=1
195     ris2=list()

```



```

196 for (i in 1:length(ris))
197   if (is.data.frame(ris[[i]]$Lz[1]))
198     {
199       ris2[[j]]=ris[[i]]
200       j=j+1
201     }
202
203 save(ris , file=paste(wd, "ris ", kk, sep=" "))
204 }
205
206 # 4 STATISTICHE SUGLI INDICI PFS
207 hit=data.frame()
208 for (kk in 1:100)
209   {
210     load(file=paste(wd, "ris ", kk, sep=" "))
211     for (i in 1:length(ris))
212       {
213         cU3=quantile(ris[[i]]$U3,.95,na.rm=TRUE)[[1]]
214         cGp=quantile(ris[[i]]$Gp,.95,na.rm=TRUE)[[1]]
215         cGpn=quantile(ris[[i]]$Gpn,.95,na.rm=TRUE)[[1]]
216         cLz=quantile(ris[[i]]$Lz,.05,na.rm=TRUE)[[1]]
217
218         nU3=length(which(ris[[i]]$U3>=cU3))
219         nGp=length(which(ris[[i]]$Gp>=cGp))
220         nGpn=length(which(ris[[i]]$Gpn>=cGpn))
221         nLz=length(which(ris[[i]]$Lz<=cLz))
222
223         hitU3=length(which((ris[[i]]$fh=="F")&(ris[[i]]$U3>=cU3)))/nU3
224         hitGp=length(which((ris[[i]]$fh=="F")&(ris[[i]]$Gp>=cGp)))/nGp
225         hitGpn=length(which((ris[[i]]$fh=="F")&(ris[[i]]$Gpn>=cGpn)))/nGpn
226         hitLz=length(which((ris[[i]]$fh=="F")&(ris[[i]]$Lz<=cLz)))/nLz
227
228         hit=rbind(hit , data.frame(hitU3, hitGp, hitGpn, hitLz , ris[[i]]$modello , ris[[i]]$percF , ris[[i]]$nitem , ris[[i]]$Q, ris[[i]]$nsogg , ris[[i]]$intF))
229       }
230     names(hit)=c("hitU3", "hitGp", "hitGpn", "hitLz", "modello", "percF", "nitem", "Q", "nsogg", "intF")
231
232     hit$percF=as.factor(hit$percF)
233     hit$Q=as.factor(hit$Q)
234     hit$nitem=as.factor(hit$nitem)
235     hit$nsogg=as.factor(hit$nsogg)
236     hit$intF=as.factor(hit$intF)
237     hit$modello=ordered(hit$modello , levels=c("extreme", "slight", "random"))
238     str(hit)
239     hit=melt(hit)
240
241     names(hit)=c("modello", "percF", "nitem", "Q", "nsogg", "intF", "indice", "value")
242
243
244     levels(hit$modello)=c("extr", "slight", "random")
245     levels(hit$indice)=c("U3", "Gp", "Gpn", "Lz")

```

```

246 hit$indice=ordered(hit$indice, levels=c("Lz", "Gp", "Gpn", "U3"))
247 levels(hit$percF)=c("percF_5%", "percF_10%", "percF_25%")
248 levels(hit$Q)=c("5_punti", "7_punti")
249 levels(hit$nititem)=c("5_item", "10_item", "20_item")
250 levels(hit$nsogg)=c("300_sogg", "600_sogg", "900_sogg")
251 levels(hit$intF)=c("p._greco_.25", "p._greco_.50", "p._greco_.75", "p._greco_.90")
252
253 save(hit, file=paste(wd, "hit", kk, sep=""))
254 }
255
256 #Carico tutti i dataset con le statistiche e li unisco
257 hit=NULL
258 for (kk in 1:100)
259 {
260     load(file=paste(wd, "hit", kk, sep=""))
261     hit2=rbind(hit2, hit)
262 }
263
264 hit=hit2
265
266 # 5 GRAFICI
267 # figura 3.1
268
269 quantile2=function(x) c(quantile(x,.05), quantile(x,.95))
270
271 load(file=paste(wd, "hit", sep=""))
272 hit2=hit[which(hit$nititem=="5_item" & hit$Q=="5_punti" & hit$nsogg=="300_sogg"),]
273
274 es=aggregate(value ~ intF + percF + indice + modello, FUN=quantile2, data=hit2)$value
275 media=aggregate(value ~ intF + percF + indice + modello, FUN=median, data=hit2)
276 new.data<-data.frame(media, lower95=es[,1], upper95=es[,2])
277
278 useOuterStrips(
279     xYplot(Cbind(value, lower95, upper95, factor=.2) ~ as.numeric(intF) | percF + modello,
280         data=new.data,
281         group=indice, type=c("l"),
282         #main="Percentuale di veri positivi nel totale di soggetti segnalati,
283         #scale da 5 item con 5 possibilita' di risposta",
284         #xlab="Probabilita' di aberrazione di una risposta (p. greco) \n
285         #cutoff fissato al 5% piu' estremo dei punteggi;
286         #percF e' la percentuale di soggetti aberrati nel campione.",
287         ylab="Modello_di_aberrazione",
288         xlab="",
289         col=c("black", "green", "red", "blue"),
290         auto.key = list(col=c("black", "green", "red", "blue"), points = FALSE, rectangles = TRUE, space = "right"),
291         par.settings=list(superpose.polygon=list(col=c("black", "green", "red", "blue"))),
292         scales=list(x=list(at=seq(1,4,1), labels = c(".25", ".50", ".75", ".90")), y=list(at=seq(0,6,0.2), labels = c("0%", "20%", "40%", "60%", "80%", "100%))),
293         labels = c("", "", "", ""), ylim = c(-.05, 1.05),
294         panel = function(x, y, ...)
295     {

```

```

326     panel.grid(h = -10, v = -3,col="grey80" ,lwd=.1, lty=3)
327     x=x+matrix(c(-.06,-.06,-.06,-.06,-.02,-.02,-.02,-.02,.02,.02,.02,.02,.06,.06,.06,.06),4,4)
328     panel.xYplot(x, y,...)
329   }
330 ))
331
332 # figura 3.2
333 hit2=hit[which(hit$nittem=="5_item" & hit$Q=="7_punti" & hit$nsogg=="300_sogg"),]
334
335 es= aggregate(value ~ intF + percF + indice + modello , FUN=quantile2, data=hit2)$value
336 media=aggregate(value ~ intF + percF + indice + modello ,FUN=median,data=hit2)
337 new.data<-data.frame(media,lower95=es[,1], upper95=es[,2])
338
339 useOuterStrips(
340   xYplot(Cbind(value, lower95, upper95, factor=.2) ~ as.numeric(intF) | percF + modello,
341     data=new.data,
342     group=indice, type=c("1"),
343     # main="Percentuale di veri positivi nel totale di soggetti segnalati,
344     # scale da 5 item con 7 possibilita' di risposta",
345     # xlab="Probabilita' di aberrazione di una risposta (p. greco) \n
346     # cutoff fissato al 5\% piu' estremo dei punteggi;
347     # percF e' la percentuale di soggetti aberrati nel campione.",
348     ylab="Modello_di_aberrazione",
349     xlab="",
350     col=c("black","green","red","blue"),
351     auto.key = list(col=c("black","green","red","blue"),points = FALSE, rectangles = TRUE, space = "right"),
352     par.settings=list(superpose.polygon=list(col=c("black","green","red","blue"))),
353     scales=list(x=list(at=seq(1,4,1),labels = c(".25",".50",".75",".90")),y=list(at=seq(0,6,0.2),labels = c("0%","20%","40%","60%","80%","100%"))),
354     labels = c("","","",""), ylim = c(-.05,1.05),
355     panel = function(x, y,...)
356     {
357       panel.grid(h = -10, v = -3,col="grey80" ,lwd=.1, lty=3)
358       x=x+matrix(c(-.06,-.06,-.06,-.06,-.02,-.02,-.02,-.02,.02,.02,.02,.02,.06,.06,.06,.06),4,4)
359       panel.xYplot(x, y,...)
360     }
361   ))
362
363 # figura 3.3
364 hit2=hit[which(hit$nittem=="20_item" & hit$Q=="5_punti" & hit$nsogg=="300_sogg"),]
365
366 es= aggregate(value ~ intF + percF + indice + modello , FUN=quantile2, data=hit2)$value
367 media=aggregate(value ~ intF + percF + indice + modello ,FUN=median,data=hit2)
368 new.data<-data.frame(media,lower95=es[,1], upper95=es[,2])
369
370 useOuterStrips(
371   xYplot(Cbind(value, lower95, upper95, factor=.2) ~ as.numeric(intF) | percF + modello,
372     data=new.data,
373     group=indice, type=c("1"),
374     # main="Percentuale di veri positivi nel totale di soggetti segnalati,
375     # scale da 20 item con 5 possibilita' di risposta",

```

```

346 #           xlab="Probabilita' di aberrazione di una risposta (p. greco) \n
347 # cutoff fissato al 5% piu' estremo dei punteggi;
348 # percF e' la percentuale di soggetti aberrati nel campione.",
349 ylab="Modello_di_aberrazione",
350 xlab="",
351 col=c("black","green","red","blue"),
352 auto.key = list(col=c("black","green","red","blue"),points = FALSE, rectangles = TRUE, space = "right"),
353 par.settings=list(superpose.polygon=list(col=c("black","green","red","blue"))),
354 scales=list(x=list(at=seq(1,4,1),labels = c(".25",".50",".75",".90")),y=list(at=seq(0,6,0.2),labels = c("0%","20%","40%","60%","80%","100%))),
355 labels = c("","","",""), ylim = c(-.05,1.05),
356 panel = function(x, y,...)
357 {
358   panel.grid(h = -10, v = -3,col="grey80" ,lwd=.1, lty=3)
359   x=x+matrix(c(-.06,-.06,-.06,-.06,-.02,-.02,-.02,-.02,.02,.02,.02,.02,.06,.06,.06,.06),4,4)
360   panel.xyplot(x, y,...)
361 }
362 ))
363
364 # figura 3.4
365 hit2=hit[which(hit$nittem=="20_item" & hit$Q=="7_punti" & hit$nsogg=="300_sogg"),]
366
367 es= aggregate(value ~ intF + percF + indice + modello , FUN=quantile2 , data=hit2)$value
368 media=aggregate(value ~ intF + percF + indice + modello ,FUN=median,data=hit2)
369 new.data<-data.frame(media,lower95=es[,1] , upper95=es[,2])
370
371 useOuterStrips(
372   xyplot(Cbind(value , lower95 , upper95 , factor=.2) ~ as.numeric(intF) | percF + modello ,
373     data=new.data ,
374     group=indice , type=c("l"),
375     #           main="Percentuale di veri positivi nel totale di soggetti segnalati ,
376     # scale da 20 item con 7 possibilita' di risposta",
377     #           xlab="Probabilita' di aberrazione di una risposta (p. greco) \n
378     # cutoff fissato al 5% piu' estremo dei punteggi;
379     # percF e' la percentuale di soggetti aberrati nel campione.",
380     ylab="Modello_di_aberrazione",
381     xlab="",
382     col=c("black","green","red","blue"),
383     auto.key = list(col=c("black","green","red","blue"),points = FALSE, rectangles = TRUE, space = "right"),
384     par.settings=list(superpose.polygon=list(col=c("black","green","red","blue"))),
385     scales=list(x=list(at=seq(1,4,1),labels = c(".25",".50",".75",".90")),y=list(at=seq(0,6,0.2),labels = c("0%","20%","40%","60%","80%","100%))),
386     labels = c("","","",""), ylim = c(-.05,1.05),
387     panel = function(x, y,...)
388     {
389       panel.grid(h = -10, v = -3,col="grey80" ,lwd=.1, lty=3)
390       x=x+matrix(c(-.06,-.06,-.06,-.06,-.02,-.02,-.02,-.02,.02,.02,.02,.02,.06,.06,.06,.06),4,4)
391       panel.xyplot(x, y,...)
392     }
393 ))

```

Bibliografia

- Beaber, R. J., Marston, A., Michelli, J., & Mills, M. J. (1985). A brief test for measuring malingering in schizophrenic individuals. *The American journal of psychiatry*.
- Burnkrant, S. R., & Harvey, R. J. (2000). Establishing baserates for the z3 and f2 appropriateness indices for use with the myers-briggs type indicator. In *annual conference of the society for industrial and organizational psychology, new orleans*.
- Campbell, A. A. (1987, MAY 29). Randomized-response technique. *Science*, *236*(4805), 1049. doi: 10.1126/science.3576215
- Caprara, G. V. (2001). *La valutazione dell'autoefficacia. Costrutti e strumenti*. (G. V. Caprara, Ed.). Trento, IT: Erikson.
- Chaudhuri, A., & Mukerjee, R. (1988). *Randomised response theory and technique*. New York: Marcel Dekker Inc.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J. E. (1987, JUN 19). Sexual-behavior and randomized responses. *Science*, *236*(4808), 1503. doi: 10.1126/science.236.4808.1503-a
- Crawford, V. P. (2003). Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *The American economic review*, *93*(1), 133–149.
- Donovan, J. J., Dwight, S. A., & Schneider, D. (2014). The impact of applicant faking on selection measures, hiring decisions, and employee performance. *Journal of Business and Psychology*. (Online First Article) doi: 10.1007/s10869-013-9318-5
- Drasgow, F. (1982). Choice of test model for appropriateness measurement. *Applied Psychological Measurement*, *6*(3), 297–308.
- Drasgow, F., Levine, M. V., & Williams, E. A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology*, *38*(1), 67–86.
- Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal Of Applied Psychology*, *84*(2), 155–166.

- Ellingson, J. E., Smith, D. B., & Sackett, P. R. (2001). Investigating the influence of social desirability on personality factor structure. *Journal Of Applied Psychology, 86*, 122–133.
- Emons, W. H. (2008). Nonparametric person-fit analysis of polytomous item scores. *Applied Psychological Measurement, 32*(3), 224–247.
- Emons, W. H., Glas, C. A., Meijer, R. R., & Sijtsma, K. (2003). Person fit in order-restricted latent class models. *Applied Psychological Measurement, 27*(6), 459–478.
- Eysenck, S. B., Eysenck, H. J., & Shaw, L. (1974). The modification of personality and lie scale scores by special 'honesty' instructions. *The British journal of social and clinical psychology, 13*, 41–50.
- Ferrando, P. J. (2005). Factor analytic procedures for assessing social desirability in binary items. *Multivariate Behavioral Research, 40*(3), 331–349.
- Ferrando, P. J., & Anguiano-Carrasco, C. (2009). Assessing the impact of faking on binary personality measures: An irt-based multiple-group factor analytic procedure. *Multivariate Behavioral Research, 44*, 497–524.
- Ferrando, P. J., & Anguiano-Carrasco, C. (2011). A structural equation model at the individual and group level for assessing faking-related change. *Structural Equation Modeling: A Multidisciplinary Journal, 18*, 91–109.
- Ferrando, P. J., & Anguiano-Carrasco, C. (2013). A structural model-based optimal person-fit procedure for identifying faking. *Educational and Psychological Measurement, 73*(2), 173–190.
- Fox, J. A., & Tracy, P. E. (1986). Randomized response: A method for sensitive surveys. In *Quantitative Applications in the Social Sciences*. SAGE Publications, Inc., California.
- Fox, J.-P., & Meijer, R. R. (2008). Using item response theory to obtain individual information from randomized response data: An application using cheating data. *Applied Psychological Measurement, 32*(8), 595–610. doi: 10.1177/0146621607312277
- Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences, 7*, 385–400.
- Galić, Z., JerneiĆ, v., & Kovačić, M. (2012). Do applicants fake their personality questionnaire responses and how successful are their attempts? a case of military pilot cadet selection. *International Journal of Selection and Assessment, 20*(2), 229–241.
- Glas, C., & Dagohey, A. V. T. (2007). A person fit test for irt models for polytomous items. *Psychometrika, 72*(2), 159–180.

- Gray, N., MacCulloch, M., Smith, J., Morris, M., & Snowden, R. (2003). Violence viewed by psychopathic murderers. *Nature*, *423*(6939), 497–498.
- Griffin, B., Hesketh, B., & Grayson, D. (2004). Applicants faking good: evidence of item bias in the neo pi-r. *Personality and Individual Differences*, *36*(7), 1545–1558.
- Griffith, R., & Converse, P. (2011). The rules of evidence and the prevalence of applicant faking. In *Faking in personality assessment: Reflections and recommendations* (pp. 34–52). Oxford University Press.
- Hall, R. C. W., & Hall, R. C. W. (2007). Detection of malingered PTSD: An overview of clinical, psychometric, and physiological assessment: Where do we stand? *Journal of Forensic Sciences*, *52*, 717–725.
- Hesketh, B., Griffin, B., & Grayson, D. (2004). Applicants faking good: evidence of item bias in the neo pi-r. *Personality And Individual Differences*, *36*, 1545–1558.
- Holden, R. R., & Book, A. S. (2009). Using hybrid rasch-latent class modeling to improve the detection of fakers on a personality inventory. *Personality And Individual Differences*, *47*(3), 185–190. doi: 10.1016/j.paid.2009.02.024
- Hopwood, C. J., Talbert, C. A., Morey, L. C., & Rogers, R. (2008). Testing the incremental utility of the negative impression-positive impression differential in detecting simulated personality assessment inventory profiles. *Journal Of Clinical Psychology*, *64*, 338–343.
- Jöreskog, K., & Sörbom, D. (1996). *Prelis 2: User's reference guide*. Lincolnwood, IL: Scientific Software International, Inc.
- Jöreskog, K., & Sörbom, D. (2001). *Lisrel 8: User's reference guide*. Lincolnwood, IL: Scientific Software International, Inc.
- Kaiser, H. F., & Dickman, K. (1962). Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix. *Psychometrika*, *27*(2), 179–182.
- Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education*, *16*(4), 277–298.
- Kelley, T. L. (1947). *Fundamentals of statistics*. Cambridge, MA: Harvard University Press.
- Kolata, G. (1987). How to ask about sex and get honest answers. *Science*, *236*(4800), 382. Retrieved from <http://www.jstor.org/stable/1698987>
- Leite, W. L., & Cooper, L. A. (2010). Detecting social desirability bias using factor mixture models. *Multivariate Behavioral Research*, *45*, 271–293.
- Levin, R. A., & Zickar, M. J. (2002). Investigating self-presentation, lies, and bullshit:

- Understanding faking and its effects on selection decisions using theory, field research, and simulation. In J. M. Brett & F. Drasgow (Eds.), *The psychology of work* (pp. 253–275). Lawrence Erlbaum Associates: Mahwah, NJ.
- Levine, M. V., & Rubin, D. B. (1979). Measuring the appropriateness of multiple-choice test scores. *Journal of Educational and Behavioral Statistics*, 4(4), 269–290.
- Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis. *Science*, 355(6325), 584–585.
- Lombardi, L., & Pastore, M. (2012). Sensitivity of fit indices to fake perturbation of ordinal data: A sample by replacement approach. *Multivariate Behavioral Research*, 47, 519–546.
- Lombardi, L., & Pastore, M. (2015). Robust evaluation of fit indices to fake-good perturbation of ordinal data. *Quality & Quantity*, Online First.
- Lombardi, L., Pastore, M., Nucci, M., & Bobbio, A. (2014). Sgr modeling of correlational effects in fake good self-report measures. *Methodology and Computing in Applied Probability*, 1–19, in press.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.
- MacCann, C., Ziegler, M., & Roberts, R. (2012). Faking in personality assessment: Reflections and recommendations. In M. Ziegler, C. MacCann, & R. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 309–329). USA: Oxford University Press.
- Marshall, E. (2000). Scientific misconduct - how prevalent is fraud? that's a million-dollar question. *Science*, 290(5497), 1662–1663. doi: 10.1126/science.290.5497.1662
- McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology*, 85, 812–821.
- Meijer, R. R. (1994). The number of guttman errors as a simple and powerful person-fit statistic. *Applied Psychological Measurement*, 18(4), 311–314.
- Meijer, R. R., Molenaar, I. W., & Sijtsma, K. (1994). Influence of test and person characteristics on nonparametric appropriateness measurement. *Applied Psychological Measurement*, 18(2), 111–120.
- Meijer, R. R., Niessen, A. S. M., & Tendeiro, J. N. (2015). A practical guide to check the consistency of item response patterns in clinical research through person-fit statistics examples and a computer program. *Assessment*, 1073191115577800.
- Meijer, R. R., & Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement*, 25(2), 107–135.
- Miller, J. D. (1981). Complexities of the randomized response solu-

- tion. *American Sociological Review*, 46(6), 928–930. Retrieved from <http://www.jstor.org/stable/2095089>
- Mittenberg, W., Patton, C., Canyock, E., & Condit, D. (2002). Base rates of malingering and symptom exaggeration. *Journal Of Clinical And Experimental Neuropsychology*, 24, 1094–1102.
- Moustaki, I., & Knott, M. (2000). Generalized latent trait models. *Psychometrika*, 65(3), 391–411.
- Muraki, E. (1993). Information functions of the generalized partial credit model. *ETS Research Report Series*, 1993(1), i–12.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent variables indicators. *Psychometrika*, 49, 115–132.
- Nešlehová, J. (2007). On rank correlation measures for non-continuous random variables. *Journal of Multivariate Analysis*, 98(3), 544 - 567.
- Pastore, M., & Lombardi, L. (2014). The impact of faking on cronbach's alpha for dichotomous and ordered rating scores. *Quality & Quantity*, 48, 1191–1211.
- Pastore, M., Nucci, M., Bobbio, A., & Lombardi, L. (2017). Empirical scenarios of fake data analysis: The sample generation by replacement (sgr) approach. *Frontiers in Psychology*, 8.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598–609.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). New York: Academic press.
- Pauls, C. A., & Crost, N. W. (2005). Effects of different instructional sets on the construct validity of the neo-pi-r. *Personality And Individual Differences*, 39(2), 297–308. doi: 10.1016/j.paid.2005.01.003
- Pearson, K. (1913). On the measurement of the influence of “broad categories” on correlation. *Biometrika*, 9(1-2), 116.
- Pek, J., & MacCallum, R. C. (2011). Sensitivity analysis in structural equation models: Cases and their influence. *multivariate behavioral research*, 46, 202–228.
- Prunas, A., Sarno, I., Preti, E., Madeddu, F., & Perugini, M. (2012). Psychometric properties of the italian version of the scl-90-r: a study on a large community sample. *European psychiatry*, 27(8), 591–597.
- R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>

- Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83*, 634–644.
- Rudner, L. M. (1983). Individual assessment accuracy. *Journal of Educational Measurement, 20*(3), 207–219.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement, 34*.
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15*(1), 72–101.
- St-Onge, C., Valois, P., Abdous, B., & Germain, S. (2009). A monte carlo study of the effect of item characteristic curve estimation on the accuracy of three person-fit statistics. *Applied Psychological Measurement, 33*(4), 307–324.
- St-Onge, C., Valois, P., Abdous, B., & Germain, S. (2011). Accuracy of person-fit statistics: A monte carlo study of the influence of aberrance rates. *Applied Psychological Measurement, 0146621610391777*.
- Topping, G. D., & O’Gorman, J. (1997). Effects of faking set on validity of the neo-ffi. *Personality and Individual Differences, 23*(1), 117–124.
- Tracy, P. E., & Fox, J. A. (1981). The validity of randomized response for sensitive measurements. *American Sociological Review, 46*(2), 187–200.
- Van der Geest, S., & Sarkodie, S. (1998). The fake patient: A research experiment in a ghanaiian hospital. *Social Science & Medicine, 47*(9), 1373–1381.
- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational And Psychological Measurement, 59*, 197–210.
- Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association, 60*(309), 63–69.
- Whitby, O. (1971). *Estimation of parameters in the generalized beta distribution* (Tech. Rep. No. 29). Stanford, CA: Department of Statistics: Standford University.
- Wilcox, R. R. (2016). *Introduction to robust estimation and hypothesis testing*. London, England: Elsevier/Academic Press.
- Zickar, M. J., Gibby, R. E., & Robie, C. (2004). Uncovering faking samples in applicant, incumbent, and experimental data sets: An application of mixed-model item response theory. *Organizational Research Methods, 7*, 168–190.
- Zickar, M. J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. *Journal of Applied Psychology, 84*, 551–563.
- Ziegler, M., & Buehner, M. (2009). Modeling socially desirable responding and its effects. *Educational And Psychological Measurement, 69*(4), 548–565.

Ziegler, M., MacCann, C., & Roberts, R. (2012). *New perspectives on faking in personality assessment*. USA: Oxford University Press.