

**UNIVERSITÀ DEGLI STUDI DI PADOVA**

**Dipartimento di Filosofia, Sociologia, Pedagogia e Psicologia Applicata  
FISPPA**

**Corso di laurea triennale in Scienze psicologiche sociali e del  
lavoro**

**SIGNIFICANZA STATISTICA E SCALA**

**LIKERT: LIMITI DI UTILIZZO**

**Meaningfulness and Likert scale: Limits of use**

*Relatore*

**Prof. Stefanutti Luca**

*Laureando: Scarpa Giacomo*

*Matricola: 1117289*

Anno Accademico 2021/2022



## ***Indice***

<b><u>INTRODUZIONE</u></b>	<b>pag. 3</b>
<b><u>CAPITOLO 1- Misurazione Fondamentale e significanza statistica</u></b>	
1.1 Cenni storici sulla misurazione in Psicologia	pag. 5
1.2 Scale di misura e invarianza statistica	pag. 7
1.3 Verso una misurazione fondamentale in Psicologia	pag. 10
<b><u>CAPITOLO 2- La scala Likert</u></b>	
2.1 Descrizione del modello	pag. 13
2.2 Equidistanza degli intervalli	pag. 16
<b><u>CAPITOLO 3- Problematiche di utilizzo di modelli metrici per variabili ordinali</u></b>	
3.1 Scala Likert: ordinale o intervallo?	pag. 19
3.2 Confutazioni	pag. 22
3.3 Errori di primo e secondo tipo	pag. 25
<b><u>CONCLUSIONE</u></b>	<b>pag. 27</b>
<b><u>APPENDICE</u></b>	<b>pag. 29</b>
<b><u>BIBLIOGRAFIA</u></b>	<b>pag. 36</b>

## **INTRODUZIONE**

Nel corso di questi anni di studio universitario spesso mi è capitato di imbattermi nella compilazione o nella somministrazione di questionari strutturati sulla base del modello Likert per la misurazione degli atteggiamenti. Infatti tale strumento di misura è uno dei più utilizzati non solo in Psicologia ma anche in tutte le scienze sociali e nei sondaggi di opinione. Nella tabella 1 in Appendice possiamo notare come negli ultimi cinque anni vi sia stato un andamento sempre crescente di pubblicazioni aventi il termine Likert come parola chiave e questo non è un dato indifferente. Per questa ragione ho provato ad osservare quanti articoli sono stati pubblicati in più da un anno all'altro negli ultimi dieci anni e i dati sono visibili nella colonna verde della tabella 2 in appendice..

In particolare trovo suggestivo il fatto che nel 2021 siano stati pubblicati 1.416 articoli in più rispetto all'anno precedente, dato sensibilmente maggiore rispetto a quelli osservati nel 2020(812), 2019(749) e a tutti gli altri compreso quello del 2013 con solo 52 articoli pubblicati in più rispetto al 2012. Questa informazione può essere interpretata anche come una conseguenza della pandemia da covid-19 poiché tale strumento di misura essendo facile da fornire anche per via telematica potrebbe essere stato utilizzato con maggiore frequenza in un periodo storico in cui risultava difficile fare ricerca a contatto diretto con i soggetti.

Frequenza e semplicità di utilizzo sono stati i motivi che mi hanno spinto ad interrogarmi sull'efficacia di tale modello del quale vi sono svariate ricerche che ne mettono in evidenza i limiti di utilizzo. Nello specifico ho deciso in questa tesi riunendo il contributo di vari autori di mettere in evidenza una problematica rilevante nell'utilizzo di questa scala di misura ovvero la violazione della significanza statistica. E' pratica sempre più comune da parte dei ricercatori adottare statistiche parametriche per il calcolo di variabili ordinali con conseguenze decisive per i risultati delle ricerche. La scala Likert è una scala ordinale secondo la classificazione di Stevens la quale però viene sempre più comunemente trattata come se fosse una scala a livello intervallo per giustificare l'adozione di modelli metrici. Tale scelta ha radici storiche profonde

che hanno portato a delle conseguenze non indifferenti e che verranno elencate nel prosieguo del lavoro.

Pertanto nel primo capitolo di matrice teorica verrà introdotto il dibattito storico sulla misurazione in Psicologia per poi soffermarsi sulla nozione di significanza statistica al fine di evidenziare la sua importanza. Il capitolo si concluderà con una considerazione sulla necessità di adottare una misurazione fondamentale nelle scienze sociali per evitare di eseguire statistiche sui punteggi piuttosto che sulle misure.

Nel secondo capitolo si descriverà la scala Likert e le sue caratteristiche mettendo in luce un primo limite di utilizzo ovvero l'impossibilità di definire equidistanti le categorie di risposta della scala.

Nell'ultimo capitolo ci si propone di mostrare come la violazione dell'invarianza statistica nel modello Likert porti a delle conseguenze rilevanti nei risultati delle ricerche. Per questa ragione verranno approfondite le tematiche riguardo alla scelta di considerare tale scala ad un determinato livello e verranno esposte le problematiche di utilizzo di statistiche parametriche nel calcolo di variabili ordinali. Tale capitolo presenta anche una parte relativa alle confutazioni di questo lavoro perchè come tutti i lavori di ricerca è doveroso riunire sempre le varie parti inserite nel dibattito scientifico dove è proprio dal dialogo tra opinioni discordanti che vengono trovate soluzioni.

Infine verranno tratte delle conclusioni relative all'intero lavoro di ricerca con la speranza che qualcuno in futuro possa proseguire nell'approfondimento di studi su tale argomento per aiutare gli esperti di settore a risolvere la controversia.

## **CAPITOLO 1 - Misurazione Fondamentale e significanza statistica**

### **1.1 Cenni storici sulla misurazione in Psicologia**

La misurazione in tutte le sue forme è stata ed è tutt'ora al centro del dibattito filosofico ed epistemologico tra le varie scienze dell'ultimo secolo. Essa nasce e si sviluppa nel contesto della fisica classica dove “una misura di una grandezza di un attributo quantitativo è una stima del rapporto tra tale grandezza e qualsiasi altra dello stesso attributo sia presa come unità di misura”<sup>1</sup>. Tale definizione mette in evidenza come la pratica della misurazione debba tenere conto necessariamente della quantificazione degli attributi degli oggetti d'indagine e questo riflette un'impostazione filosofica di stampo Euclideo dove qualsiasi attributo deve essere descritto come multiplo di una grandezza standard ovvero l'unità di misura così da permetterne una concatenazione.

Fin dagli albori della sua nascita la psicologia di fronte al dilemma riguardo la possibilità di eseguire stime quantitative delle sensazioni umane si è sempre dovuta confrontare con il grande spauracchio della misurazione. La questione si pone soprattutto sul piano ontologico poiché nelle scienze esatte gli oggetti di indagine possiedono delle caratteristiche manifeste, manipolabili ed osservabili diversamente dalle scienze sociali dove lo studio delle variabili si pone sul piano latente alimentando così il dibattito sulla definizione e costruzione degli strumenti di misura.

Per uscire da questo stallo gli scienziati di fine Ottocento adottarono le tecniche della psicofisica classica per la misurazione dei tratti mentali (per esempio Weber-Fechner con soglie e metodi indiretti, Thurstone con signal detection theory) le quali evidenziavano il fatto che la realtà non sempre coincide con la percezione della stessa poiché i sensi sono ingannevoli e l'unico modo per uscire dall'inganno è la conoscenza ottenibile grazie ad una rigorosa misurazione. Tutto questo ha permesso che nel 1938 la British Association for

---

<sup>1</sup> Michell J., *Representational theory of measurement. In Marcel Boumans (Eds.), Measurement in Economics: A Handbook*, 2007, Elsevier Science, Amsterdam.

the Advancement of Science si interrogasse in via ufficiale sulla possibilità o meno di eseguire misurazioni nelle scienze comportamentali e psicologiche.

Secondo Holder O. L. (1901), il processo di misurazione riguarda attributi di un dato oggetto di studio che devono avere una struttura quantitativa ovvero devono poter soddisfare determinati requisiti. Il primo riguarda i gradi delle caratteristiche che tali attributi posseggono infatti ogni ente misurato deve possedere la stessa caratteristica purché sia in grado diverso. In secondo luogo i destinatari della misurazione devono poter essere ordinati totalmente per quel loro specifico attributo per cui non esiste misura senza ordinamento delle caratteristiche degli oggetti. Infine l'ultima prerogativa consiste nell'obbligo di poter fare operazioni additive tra gli oggetti di studio. Facendo un esempio possiamo dire che la misura della massa di un mattone possiede tutte le caratteristiche quantitative sopra elencate poiché tutti i mattoni possiedono l'attributo massa, sono ordinabili grazie ad essa ed è possibile eseguire operazioni additive tra le masse dei vari mattoni. Questi dettami si basavano su una concezione della misurazione in linea con quella delle scienze esatte e non risultavano efficaci per la psicologia la quale con gli strumenti dell'epoca non riuscì a dimostrare la struttura quantitativa delle proprie variabili. Nel rapporto finale della commissione si legge: "Qualsiasi legge che si proponga di esprimere una relazione quantitativa tra l'intensità della sensazione e dello stimolo non è semplicemente falsa ma è, di fatto, priva di significato salvo che e fintanto che non possa essere dato un significato al concetto di addizione applicato alla sensazione"<sup>2</sup>.

Ho trovato molto interessante la risposta di Stevens il quale replica al commento sopra riportato notando che "il vero problema è il significato della misurazione"<sup>3</sup> e che "forse l'accordo può essere raggiunto meglio se riconosciamo che la misurazione esiste in una varietà di forme e queste scale di misura rientrano in determinate classi definite"<sup>4</sup>. Partendo da questo presupposto egli propone una nuova definizione di misura come "assegnazione

---

<sup>2</sup> Ferguson et al., *Quantitative estimates of sensory events: final report*, 1940, Advancement of science, p.245.

<sup>3</sup> Stevens S. S., *On the theory of scales of measurement*, 1946, Science, p. 677.

<sup>4</sup> Ibidem.

di numeri a oggetti o eventi secondo delle regole"<sup>5</sup>. Vorrei porre l'attenzione sul fatto che egli non utilizza il termine numbers che in inglese indica i numeri intesi come valori espressivi di una proprietà bensì il termine numerals ovvero numeri come significanti per giustificare l'utilizzo del numero anche solo come etichetta qualitativa.

## 1.2 Scale di misura e invarianza statistica

Stevens continua nella sua disamina evidenziando come misurare, quindi costruire una scala di misura, significa di fatto stabilire una terna tra un sistema relazionale empirico, un modello formale (sistema numerico) ed un morfismo. Pertanto due scale sono equivalenti quando riescono a misurare lo stesso sistema empirico attraverso una trasposizione numerica che ci consenta di passare da una scala all'altra ovvero una trasformazione permissibile. A questo punto vengono elaborate quattro condizioni metodologiche differenti dove i numeri rimandano ad un diverso significato in base a quello che essi vogliono esprimere. Per fare questo si pone come condizione necessaria e sufficiente il rispetto dei due criteri di rappresentazione e unicità dove nel primo caso si "provvede all'assegnazione di numeri a oggetti del mondo empirico"<sup>6</sup> e nel secondo si indicano "le tipologie di classi o funzioni con cui si devono relazionare le due rappresentazioni numeriche"<sup>7</sup>. In particolare il teorema di unicità serve ad andare contro il problema dell'indeterminazione poiché dati sistemi relazionali empirici e morfismi tramite le trasformazioni permissibili si possono misurare due oggetti in maniera diversa. Conseguenza di tutto ciò è stato il delinearsi di quattro scale di misura e delle operazioni statistiche applicabili ai vari livelli di ogni scala.

La prima scala proposta da Stevens è la nominale dove "i numeri sono utilizzati come etichette o simboli"<sup>8</sup> per classificare gli oggetti e quindi diventano di fatto delle etichette verbali per creare classi di equivalenza. Questo tipo di scala ha

---

<sup>5</sup> Stevens S. S., *On the theory of scales of measurement*, 1946, Science, p. 678

<sup>6</sup> Townsend J. T. and Ashby F. G., *Measurement scales and statistics: The misconception misconceived*. Psychological Bulletin, 1984, p. 395

<sup>7</sup> Ibidem.

<sup>8</sup> Stevens S. S., *On the theory of scales of measurement*, 1946, Science, p. 678.



subìto delle forti critiche, per esempio Lord F. afferma che “[Secondo Stevens,] l’unica regola non ammessa sarebbe l’assegnazione casuale. Tuttavia, la cosiddetta misurazione nominale comporta assegnazioni arbitrarie”<sup>9</sup>. Sta di fatto che per la definizione di misurazione che aveva precedentemente dato Stevens tale tipo di scala rientra all’interno delle modalità di misura poiché i numeri possono essere anche considerati solo mere etichette. L’unicità di tale scala è garantita tramite l’utilizzo di una funzione iniettiva o biunivoca.

La seconda tipologia di scala è la ordinale dove “una proprietà è presente in un oggetto più che in un altro”<sup>10</sup> senza fornire indicazioni sulla distanza tra gli elementi si limita solo a definire l’ordine la cui invarianza è possibile solo all’interno di una funzione monotona crescente in senso stretto (sempre crescente nel dominio).

La possibilità di introdurre un’unità misura arriva con la scala intervallo che secondo Stevens è “quantitativa nel senso ordinario del termine”<sup>11</sup> poiché permette di stabilire le differenze tra i vari elementi misurati anche se non in maniera assoluta poiché è una scala arbitraria la cui trasformazione permessa è una funzione lineare positiva.

L’ultima è la scala rapporto che attraverso la similitudine diretta permette di “determinare tutte e quattro le relazioni: uguaglianza, ordinamento, uguaglianza degli intervalli e uguaglianza dei rapporti”<sup>12</sup>; questo tipo di scale come fa notare Coombs “sono la norma nelle scienze fisiche e nella vita di tutti i giorni”<sup>13</sup>.

A questo punto il problema diventa capire quali sono “le operazioni statistiche applicabili alle misurazioni realizzate con ogni tipo di scala”<sup>14</sup> e quindi di fatto chiedersi se una determinata statistica è significativa a livello di scala. Interrogarsi sulla significanza di una statistica significa chiedersi se essa è invariante secondo una delle tre forme di varianza (assoluta, riferimento e confronto) poiché se una statistica non è significativa vuol dire che non ha senso applicare quella statistica a livello di quella scala altrimenti sarebbe una

---

<sup>9</sup> Lord, F.M. *On the statistical treatment of football numbers*. *American Psychologist*, 1953, p.750

<sup>10</sup> Bruschi A., *Metodologia delle scienze sociali*, 1999, Milano, Bruno Mondadori, p. 82.

<sup>11</sup> Stevens S. S., *On the theory of scales of measurement*, 1946, Science, p. 679.

<sup>12</sup> Ibidem.

<sup>13</sup> Coombs. C H. [1950] *Psychological Scaling Without a Unit of Measurement*, 1950 *Psychological Review*, vol. LVII, n. 3, pp. 145-58

<sup>14</sup> Ivi p. 677.

statistica dubbia. Nella tabella 3 in appendice si possono osservare nell'ultima colonna tutte le statistiche significanti per ogni specifica scala in relazione con le trasformazioni caratteristiche della stessa secondo il principio di invarianza statistica. Vorrei in particolare soffermarmi sulla media che come facilmente dimostrabile è una statistica invariante a livello di scala intervallo ma non a livello di scala ordinale e questo verrà rimarcato anche nel successivo capitolo. Citando Roberts si potrebbe definire il concetto di invarianza nei seguenti termini: "un'affermazione che utilizza scale è chiamata significativa se la sua verità o falsità rimane invariata quando tutte le scale nell'enunciato vengono trasformate da trasformazioni ammissibili"<sup>15</sup> pertanto "la statistica non è scala-dipendente"<sup>16</sup>. Sebbene questo aspetto venga accettato per le operazioni di statistica descrittiva non si può dire lo stesso quando l'argomentazione volge nell'ambito dell'inferenza statistica. Alcuni autori sostengono che l'accuratezza di un test di ipotesi è indipendente dalle caratteristiche della scala di misura purché le procedure di campionamento e il modello matematico di inferenza vengano rispettati. Gaito per esprimere questo concetto nel suo famoso articolo sul *Psychological Bulletin* cita Lord che dice che "i numeri non si ricordano da dove vengono"<sup>17</sup>; una frase sicuramente di impatto ma a mio avviso pericolosa perché i numeri sono strumenti che noi utilizziamo per descrivere un mondo empirico e sta a noi decidere le regole per cui utilizzarli e che riflettano realisticamente le dinamiche di tale mondo. Sicuramente i test di ipotesi dipendono dal campionamento e dalle procedure utilizzate per condurre tale test ma è chiaro che i dati che io rilevo da quel campione sono calcolati secondo una certa statistica a livello di una determinata scala e se tale statistica non è significativa anche se il t-test è valido non è significativa. Come dice Roberts: "Se verifichiamo un'ipotesi su una popolazione sconosciuta vogliamo che l'ipotesi dica qualcosa di significativo sulla popolazione non qualcosa che sia un artefatto della particolare scala di misura utilizzata"<sup>18</sup>. Quindi il centro di questo dibattito non sta tanto nel test di verifica dell'ipotesi quanto nell'ipotesi stessa che deve poter garantire i criteri di rappresentazione e unicità propri

---

<sup>15</sup> Roberts M. et al., *Meaningless Statistics*, 1987, *Journal of Education statistics*, pp. 383-384

<sup>16</sup> *Ibidem*

<sup>17</sup> Lord, F.M. *On the statistical treatment of football numbers. American Psychologist*, 1953, p. 751

<sup>18</sup> Roberts M. et al., *Meaningless Statistics*, 1987, *Journal of Education statistics*, p. 390

della misurazione fondamentale. Potremmo dire che i numeri non si ricordano da dove vengono ma i ricercatori invece sì e decidono come utilizzarli.

### **1.3 Verso una misurazione fondamentale in Psicologia**

L'obiettivo principale di questo lavoro compilativo è interrogarmi sulle problematiche di impiego della scala Likert per la misurazione dei costrutti psicologici in virtù dei limiti che le scale di misura impongono all'utilizzo di alcune statistiche ed uno degli aspetti che vorrei mettere in luce riguarda l'utilizzo inappropriato di alcune statistiche calcolate su punteggi grezzi i quali vengono trattati come se fossero misure continue. "L'articolazione in livelli di scala proposta da Stevens viene usata come pretesto da molti ricercatori che sono soliti costruire i punteggi contando il numero di risposte corrette date ad un questionario e trattare tali punteggi come se fossero statistiche sufficientemente buone per ottenere una misurazione valida"<sup>19</sup>.

Nel volume "Measurement in psychology: a critical history of a methodological concept" Michell(1999) spiega che gli psicologi hanno preferito crearsi un personale metodo di misurazione stipulando delle convenzioni formali come quella di Stevens sui livelli di scala piuttosto che utilizzare il metodo di misurazione comune a tutte le scienze diventando in un certo senso autoreferenziali e pseudoscientifici. Personalmente sono d'accordo con questa critica e ritengo che fare operazioni aritmetiche sui punteggi grezzi come fossero misure influenzi negativamente la qualità e la credibilità delle ricerche psicologiche anche perché come dice Reese: " misurazione é semplicemente una parola usata per indicare certe idee. Usarla per indicarne altre non amplia il suo significato ma lo distrugge"<sup>20</sup>.

Quello a cui le scienze sociali devono ambire è la misurazione fondamentale ovvero la misurazione che non deriva da altre misurazioni e nella quale siano

---

<sup>19</sup> Robusto.E. *La misurazione fondamentale in psicologia in la sfida della misurazione nelle scienze sociali a cura di Paolo di Nicola*,2012,Franco Angeli s.r.l. p.36

<sup>20</sup> Reese T.W. *The application of the theory of physical measurement to the measurement of psychological magnitudes,with three experimental examples*,1943,Psychological Monographs,55:1-89

evidenti tratti quantificabili ed unità di misura dell'oggetto di indagine. Essa è stata introdotta nelle scienze sociali da Luce e Turkey i quali diversamente dalla maggior parte dei ricercatori ritenevano che non fosse un'utopia trattare le caratteristiche psicologiche dei soggetti in termini di quantità. Anche Ramsay(1975) ha sottolineato l'importanza di produrre una misurazione fondamentale per lo sviluppo del sapere scientifico: "Potremmo dover attendere la misurazione fondamentale prima di vedere un qualsiasi progresso nelle leggi quantitative del comportamento. In breve le scale ordinali forse non sono abbastanza buone e potrebbe non essere possibile vivere per sempre con una dozzina di procedure diverse per quantificare lo stesso pezzo di comportamento, ciascuna delle quali fa assunzioni forti ma non verificabili e fundamentalmente improbabili che portano ad intrecci non lineari di una scala contro un'altra. I progressi nel campo della fisica sarebbero stati incredibilmente difficili senza la misurazione fondamentale, ed il lettore che crede che tutto ciò che è in gioco nel trattamento assiomatico della misurazione sia una possibile canonizzazione di una procedura di scaling a scapito delle altre sta mancando il punto. Una razionalizzazione della quantificazione può essere un presupposto necessario della Psicologia come Scienza Quantitativa Razionale"<sup>21</sup>. Effettivamente anche il termometro inizialmente non teneva conto della pressione atmosferica perché non era isolato ma non è stata modificata la definizione di misura per giustificare lo strumento piuttosto nel tempo si è cambiato lo strumento rendendo la misurazione più precisa e coerente: uno strumento di misura che cambia a seconda di ciò che deve misurare non è uno strumento valido.

In questo capitolo abbiamo introdotto storicamente il dibattito sulla misurazione arrivando alla formulazione delle scale di misura di Stevens per poi evidenziare la nozione di significanza delle statistiche. Questo ci ha portati a compiere una riflessione sulla misurazione fondamentale ed è emerso che nell'ambito della ricerca psicologica si è soliti trattare i punteggi come fossero misure quantificabili cadendo così in una serie contraddizioni di natura epistemologica che si riflettono sulla qualità delle ricerche. Nel prossimo capitolo mi soffermerò

---

<sup>21</sup> Ramsay J.O., *Review of foundation of measurement*, in Psychometrika by rantz et al., 1975 Vol I, 40: p 262

nel descrivere un particolare tipo di scala frequentemente utilizzata nella ricerca sociale ovvero la scala Likert mettendone in evidenza i limiti di utilizzo alla luce della nozione di significanza statistica.

## **CAPITOLO 2- La scala Likert**

### **2.1 Descrizione del modello**

Uno dei dibattiti che ha maggiormente caratterizzato gli psicologi interessati allo studio della personalità del primo Novecento riguarda la struttura degli atteggiamenti in particolare sul fatto che essi possano essere descritti da una prospettiva disposizionale piuttosto che situazionale. Nel primo caso si considerano i tratti di personalità come basi genetiche stabili nel tempo e nello spazio, presenti fin dalla nascita ed indipendenti dal contesto sociale in cui si inseriscono diversamente dal secondo dove l'ambiente circostante coi suoi numerosi stimoli viene visto come la causa principale dell'atteggiamento predittore del comportamento umano. Scegliere se si vuole studiare la personalità secondo il criterio di specificità o generalità significa decidere se si ritiene più opportuno andare a ricercare i tratti latenti di un soggetto attraverso uno strumento di misura o lavorare sugli stimoli esterni per influenzare un comportamento. Rensis Likert enuclea il problema sulla specificità o generalità degli atteggiamenti quando dice: "Sia che prendiamo l'atteggiamento come entità innata o appresa, in entrambi i casi non è un elemento inflessibile e rigido della personalità ma piuttosto un certo intervallo entro il quale si muovono le risposte. Su questa base uno dei nostri problemi cardinali è trovare se gli atteggiamenti sociali, in questo senso, possono dimostrarsi misurabili e se arriva una risposta affermativa un serio tentativo deve essere fatto per giustificare la separazione di un atteggiamento da altri. Perché se non ci sono differenze familiari tra gli atteggiamenti, c'è semplicemente una serie infinita di atteggiamenti"<sup>22</sup>. Il fatto che le risposte si muovono entro un certo intervallo fa implicitamente intendere che Likert considera l'atteggiamento come una misura continua nonostante poi lo strumento che viene elaborato successivamente abbia una struttura discreta.

---

<sup>22</sup> Likert, R. *A Technique for the Measurement of Attitudes*. Archives of Psychology, 1932, monograph no. 140, p 8

Come abbiamo potuto notare non è poi così semplice definire cosa sia un atteggiamento, quale sia la sua struttura e soprattutto come sia possibile misurarne gli attributi. La definizione di un costrutto psicologico è un'operazione complessa e noi in questa sede riportiamo la scelta di Park di descrivere gli atteggiamenti come "modi preferenziali di comportarsi/reagire in una specifica circostanza radicata relativamente ad un'organizzazione duratura di credenze e idee (intorno ad un oggetto, un soggetto o un concetto) acquisiti attraverso le interazioni sociali"<sup>23</sup>. Se già di per sé la definizione di un atteggiamento è un processo lungo ancor di più ciò è valido per la costruzione di uno strumento di misura infatti l'obiettivo di Rensis Likert era proprio quello di elaborare una scala adatta a "misurare gli atteggiamenti in un modo scientificamente accettato e convalidato"<sup>24</sup>.

La tecnica consiste nel mettere a punto una serie di item ritenuti indicatori di un atteggiamento favorevole o contrario verso un dato argomento chiedendo poi ai soggetti di indicare su una scala bipolare (il più delle volte a 5 o 7 punti) il grado di accordo e disaccordo con la frase presente nell'item. In letteratura si è spesso dibattuto riguardo la "migliore usabilità possibile in termini di affidabilità e validità del numero di punti della scala"<sup>25</sup>. Una buona parte di esperti di settore ritiene che a maggiori categorie di risposta corrispondano misurazioni più precise in particolare "la scala a 7 punti fornisce più varietà di opzioni che a loro volta aumentano la probabilità di incontrare la realtà oggettiva delle persone"<sup>26</sup>.

In un interessante studio simulativo due ricercatori H.Wu & S.Leung analizzano la possibilità di aumentare il numero di punti della scala a 11 concludendo che "aumentare il numero di punti della scala Likert farà avvicinare la scala alle distribuzioni sottostanti"<sup>27</sup> in modo tale che come dice Alwin "più

---

<sup>23</sup> Park K. *Social sciences and health*. In Park K, editor. Park Text Book of Preventive and Social Medicine. Jabalpur: Bhanot Publishers. 2013;54x-54x.)

<sup>24</sup>A.Joshi et al. Likert scale:Explored and explained .British journal of applied science & technology 2015 p397

<sup>25</sup> "Colman AM, et al *Comparing rating scales of different lengths: Equivalence of scores from 5-point and 7-point scales*. Psychological Reports. 1997;80:355-362

<sup>26</sup> Ibidem

<sup>27</sup> H.Wu,S.Leung,Can Likert Scales be treated as Interval Scales-A simulation study,Journal of Social Service Research,2017 p 5

informazioni verranno trasmesse con più punti e quindi più categorie, più i risultati misureranno in modo più preciso”<sup>28</sup>.

Per l'attribuzione dei punteggi avendo la scala un livello totalmente ordinato si assegna semplicemente un numero ad ogni categoria di risposta mantenendo intatto l'ordine. Il punteggio totale osservato non è altro che la somma dei punteggi dati dal soggetto ad ogni item in una scala di tipo additivo. Il modello Likert si basa su una costruzione di una scala multi-item dove si presume che gli errori casuali di misura si annullino nella media pertanto il punteggio sommato sarà più attendibile.

Il nocciolo della questione però non è tanto la costruzione dello strumento di misura quanto piuttosto l'analisi dei dati che viene fatta in seguito. In questa sede ci si vuole interrogare sulla significanza delle statistiche applicate alla scala Likert quindi su quanto tale scala di misura impone dei limiti ai parametri che si vogliono misurare e soprattutto se tali parametri che vengono calcolati sono effettivamente invarianti. Questo studio ha la sua origine e il suo termine attorno ad una scelta fondamentale che deve fare ogni ricercatore ovvero se considerare la scala Likert come una scala ordinale o intervallo secondo le categorie proposte da Stevens(vedi sopra). Decidere di seguire una scuola di pensiero piuttosto che un'altra vuol dire analizzare gli item in maniera completamente diversa e trarre quindi conclusioni differenti. Likert nella sua opera monografica “A Technique for the Measurement of Attitudes” del 1932 non si azzarda a inserire la sua scala nell'orizzonte della cardinalità ma la reintroduce nell'analisi delle valutazioni sommate(Summated ratings). Utilizzare infatti una scala di tipo additivo significa dire implicitamente che esistono degli intervalli tra i punti della scala altrimenti non si potrebbero sommare(5+6 fa 11 ma quinto più sesto non fa undicesimo).

---

<sup>28</sup> Alwin. *Feeling thermometers versus 7-point scales*.  
Sociological Methods and Research, 1997, 3(25), 318–340



## 2.2 Equidistanza degli intervalli

L'adozione di un sistema classificatorio nella costruzione di una scala impone l'ordinamento degli elementi senza specificarne la distanza. Analizzare i dati di una scala Likert a livello ordinale equivale a dire che la distanza tra i vari punti della scala non è definita in quanto l'unico fattore determinante è l'ordine di tali punti. In questo senso associare il numero 7 all'affermazione "molto d'accordo" ed il numero 6 ad "abbastanza d'accordo" significa dire che la prima affermazione è più grande della seconda ma non è possibile specificare di quanto sia più grande poiché siamo ancora a livello ordinale. Le statistiche che successivamente si potranno calcolare sulla batteria di test sono quelle indicate nella figura 1 in appendice secondo la classificazione di Stevens poiché sono le uniche statistiche invarianti a questo livello di scala pertanto sarà possibile trovare il numero delle classi di equivalenza, il percentile, la moda e la mediana.

Nel momento in cui un ricercatore decide di eseguire altre statistiche (es. la media) significa che sta calcolando i dati a livello di scala intervallo poiché altrimenti starebbe utilizzando un parametro non significativo. Tale scelta parte da un presupposto fondamentale ovvero quello di considerare equivalenti le categorie di tale scala infatti come dice Joshi "Il conflitto nasce facendo un'altra domanda: i punti in scala sono equidistanti? (in altre parole è "neutro" alla stessa distanza da "d'accordo" come "d'accordo" da "fortemente d'accordo)?"<sup>29</sup>.

Nehemiah Jordan nel suo articolo "The Asymmetry of "Liking" and "Disliking": A Phenomenon Meriting Further Reflection and Research" fu il primo a contestare l'equidistanza delle categorie di una scala Likert in particolare della equidistanza delle ali dal centro in quanto a suo avviso manifestare il proprio disaccordo è psicologicamente più impegnativo che dichiararsi in accordo. Benzécri, nel secondo volume della sua *Analyse des données* elaborò la tecnica dell'*analisi delle corrispondenze* (ac) la quale può essere usata anche per "stimare la distanza mediamente percepita da un campione di intervistati fra

---

<sup>29</sup> A. Joshi et al. Likert scale: Explored and explained. British journal of applied science & technology 2015 p398

le posizioni di una scala”<sup>30</sup>. “ L’ac è una tecnica di tipo fattoriale che fornisce rappresentazioni sintetiche di vaste matrici di dati. Come tutte le tecniche di tipo fattoriale, sintetizza le variabili attraverso una o più combinazioni delle stesse, che vengono chiamate fattori, e raggruppa i casi che si presentano omogenei rispetto a un certo gruppo di variabili”<sup>31</sup>. L’aspetto interessante di questa tecnica è che raggruppando insieme matrici di dati ne dà una rappresentazione geometrica immediata ed evidente grazie alla quale “la vicinanza tra i punti è interpretata come prossimità semantica”<sup>32</sup>. Nella Figura 1 in appendice possiamo osservare come dovrebbe essere il diagramma delle ac se le categorie di una scala Likert fossero equidistanti..

Un interessante lavoro di ricerca pubblicato da Marradi<sup>33</sup> utilizza la tecnica dell’ac per confrontare dodici batterie di test diversi tra loro con l’intento di verificare la presunta equidistanza dei punti della scala. Nella Tabella 4 in Appendice sono visibili i questionari analizzati mentre dalla figura 2 alla 12 i vari grafici risultanti dall’analisi dei dati i quali vanno confrontati con la figura 1 dove si può notare come dovrebbe essere una rappresentazione coerente con l’assunto di equidistanza delle categorie. Nella visualizzazione geometrica dei grafici è importante considerare che nell’asse delle ascisse si rappresenta il fattore accordo-disaccordo mentre nell’asse delle ordinate si valuta l’intensità dell’atteggiamento sicchè a valori negativi dell’ordinata corrispondono risposte considerate poco impegnative. I risultati di questo studio vengono divisi in due gruppi dove nel primo ci sono quelli che sembrano in parte corroborare l’assunto di equidistanza degli intervalli mentre nel secondo sono presenti quelli che confutano tale principio anche se gli autori specificano che “da nessuna analisi è peraltro emersa un’equidistanza davvero soddisfacente tra le categorie di risposta proposte dalle scale Likert.”<sup>34</sup>

---

<sup>30</sup> Benzécri J. P. (1973; II ed. 1982), *L’analyse des données*, vol. II, *L’analyse des correspondances*, Paris: Dunod. 475 ss

<sup>31</sup> Di Franco G., Marradi A., *Analisi dei fattori e analisi in componenti principali*, Acireale-Roma:2003 Bonanno.

<sup>32</sup> Amaturò E., *Analyse des données e analisi dei dati nelle scienze sociali*, Torino:1989 Centro Scientifico.

<sup>33</sup> Marradi A., Gasperoni G., *Costruire il dato 3. Le scale Likert*, Milano:2002 Franco Angeli

<sup>34</sup> Marradi A. Macri.E *Sono equidistanti le categorie di una scala Likert?* Cambio:rivista sulle trasformazioni sociali,Firenze,2012,p 177

Particolarmente curioso è il fatto che vi siano addirittura due casi in cui l'ordine delle categorie di risposta non viene rispettato come possiamo notare nelle figure 11 e 12 il che potrebbe suggerire che in alcuni casi la scala Likert non soddisfa neanche il livello ordinale. I ricercatori osservano che gli intervistati discriminano bene solo fra due categorie del disaccordo per cui sono preferibili scale a 5 piuttosto che a 7 posizioni in controtendenza con ciò che sosteneva Hodge per cui "aumentare il numero di punti porterà la scala più vicino ad una misura continua"<sup>35</sup>. Inoltre la categoria intermedia si pone più vicino alle categorie del disaccordo piuttosto che a quelle dell'accordo per questo al termine delle analisi si può concludere che per quanto i "controlli eseguiti e descritti non sono così univocamente negativi per l'assunto di equidistanza delle categorie usate nelle scale Likert"<sup>36</sup> rimane il fatto che "non si può affatto dare per scontato che le posizioni di una scala Likert siano percepite come più o meno equidistanti: il più delle volte non è così"<sup>37</sup>.

Il lavoro di ricerca sopra riportato ci fa capire un aspetto importante ovvero che i presupposti che tante volte diamo per scontati nella costruzione di uno strumento di misura spesso possono essere contraddittori. L'impossibilità di dimostrare scientificamente l'equidistanza tra le categorie di risposta del modello Likert ci mette nella condizione di non poter eseguire statistiche a livello di scala intervallo per esso. Non è un caso che l'esecuzione di statistiche parametriche necessiti di condizioni minime che sono "la randomizzazione dei dati, l'indipendenza dei dati, l'uguaglianza della varianza e gli intervalli minimi"<sup>38</sup>. La scala intervallo infatti si caratterizza per la possibilità di introdurre un'unità di misura che ne garantisce l'invarianza dei confronti ( $190-185=10-5$ ) ma se i punti della scala hanno distanze diverse e quindi anche intervalli minimi differenti tale invarianza non è garantita.

---

<sup>35</sup> Hodge, D. R., & Gillespie, D. F.. Phrase completion scales: A better measurement approach than Likert Scales?. *Journal of Social Service Research*, 2007, 33(4), 1–12.

<sup>36</sup> Marradi A. Macri. *E Sono equidistanti le categorie di una scala Likert?* Cambio: rivista sulle trasformazioni sociali, Firenze, 2012, p 186

<sup>37</sup> Ibidem

<sup>38</sup> Pell.G. *Use and misuse of Likert scales*, Medical Education, 2005, p 970

## **CAPITOLO 3- Problematiche di utilizzo di modelli metrici per variabili ordinali**

### **3.1 Scala Likert: ordinale o intervallo?**

All'inizio di questo lavoro compilativo si è introdotto il tema della misurazione nelle scienze sociali delineando in particolar modo l'importanza della significanza statistica all'interno della teoria sui livelli di scala proposta da Stevens. Successivamente è stata presa in considerazione un particolare tipo di scala ( la scala Likert ) ed è stata evidenziata l'impossibilità di garantire un'equidistanza tra le categorie di risposta della stessa. La mancata equivalenza degli intervalli della scala costringe il ricercatore a trattare tale tipo di scala a livello ordinale impedendogli di poter applicare statistiche parametriche in quanto violerebbe il principio di significanza statistica. Diversamente da quanto ci si aspetterebbe però si assiste sempre più spesso all'utilizzo di modelli metrici nel calcolo di variabili ordinali presenti in tale tipo di scala e questo porta a delle problematiche non indifferenti. Nel 2016 Liddell e i suoi collaboratori analizzarono le pubblicazioni di quell'anno di tre riviste scientifiche rinomate a livello accademico (Journal of Personality and Social Psychology (JPSP), Psychological Science (PS), and the Journal of Experimental Psychology:General (JEP:G)) prendendo in considerazione tutte le ricerche che utilizzavano il modello Likert per osservare a quante di queste venissero applicate statistiche parametriche. Su 78 articoli visionati 10 non vennero considerati poiché non utilizzavano le variabili Likert come variabili dipendenti o erano ricerche bibliografiche o citazioni mentre per i restanti 68 articoli conclusero che "ognuno ha trattato i dati ordinali come metrici e ha utilizzato un modello metrico; non un'analisi singola nei 68 articoli utilizzava un modello ordinale"<sup>39</sup>. Questa ricerca mette in risalto come sia pratica sempre più comune trattare la scala Likert a livello intervallo anziché a livello ordinale come dovrebbe essere.

---

<sup>39</sup>M. Liddell et al., *Analyzing ordinal data with metric models: What could possibly go wrong?* Department of Psychological and Brain Sciences, Indiana University, 2018 p.329

“Le chiavi di questo aspetto della controversia sono le nozioni di appropriatezza e significanza”<sup>40</sup>. Come sottolineano Marcus-Roberts e Roberts “è sempre opportuno calcolare le medie (ad esempio) per le scale ordinali, ma non è opportuno fare certe affermazioni su tali medie”<sup>41</sup>. Nello specifico è importante ricordare che un'affermazione significativa non è per forza vera ma la sua falsità o verità non dipende dalla scala di misura poiché come dice Roberts “sono alto il doppio della Sears Tower. Questa affermazione è significativa, poiché la sua verità o falsità è invariante rispetto a qualsiasi trasformazione ammissibile, ma è ovviamente falsa per tutte le scale che misurano l'altezza”<sup>42</sup>. In realtà lo stesso Stevens nella formulazione della sua teoria dava una certa ammissibilità all'interpretazione più flessibile della significanza statistica scrivendo: “Nella più rigorosa correttezza le statistiche che coinvolgono medie e deviazioni standard non dovrebbero essere utilizzate con queste scale (ordinali), poiché queste statistiche implicano la conoscenza di qualcosa di più della relativa graduatoria dei dati. D'altra parte, per questa statistica 'illegale' si può invocare una sorta di sanzione pragmatica: numerosi casi portano a risultati fruttuosi”<sup>43</sup>. Ciò non toglie che la scienza dovrebbe procedere non per visioni pragmatiche flessibili o inflessibili quanto piuttosto da regole condivise che rispettino l'epistemologia e il principio di falsificazione per cui “una teoria è scientifica solo se è in grado di suggerire quali esperimenti e osservazioni potrebbero dimostrarla falsa”<sup>44</sup>.

Kuzon nel suo articolo “The seven deadly Sins of statistical analysis” attacca i sostenitori di una visione pragmatica dell'analisi statistica annoverando come primo peccato capitale proprio l'adozione di statistiche parametriche per variabili ordinali. L'autore scrive che indipendentemente dal disegno della ricerca i parametri sono sempre frutto di un calcolo per cui “moltiplicazione e divisione sono utilizzati per calcolare media e varianza”<sup>45</sup>. In letteratura vi sono alcuni

---

<sup>40</sup> Knapp TR. *Treating ordinal scales as interval scales: an attempt to resolve the controversy*. Nurs Res 1990;122

<sup>41</sup> Roberts M. et al. *Meaningless Statistics*, Journal of Education statistics, 1987, 12 p 394

<sup>42</sup> Ibidem

<sup>43</sup> Stevens S. S., *On the theory of scales of measurement*, 1946, Science, p. 680.

<sup>44</sup> Fuso.S., *Realtà o illusione? Scienza, pseudoscienza e paranormale*, Bari, Edizioni Dedalo, 1999.

<sup>45</sup> Kuzon et al. *The seven deadly sins of statistical analysis*. Ann Plast surg, University of Michigan 1996; 37:265–72.

autori che hanno provato a trovare un punto di incontro tra la visione più radicale e quella più pragmatica dell'analisi statistica per variabili ordinali per esempio Knapp che esaminando il lavoro di svariati ricercatori ha proposto alcuni accorgimenti per permettere di conciliare le due fazioni. In particolare egli si sofferma sulla necessità di eseguire un campionamento accurato ed impostare un disegno della ricerca coerente con ciò che si vuole misurare dicendo: "quando stai costruendo la tua scala (operazionalizzando il tuo costrutto, scegliendo le tue categorie, ecc.) affronta onestamente la tassonomia di Stevens e fissati a un livello particolare"<sup>46</sup>. Diciamo che sicuramente il campionamento, le distribuzioni e il disegno della ricerca sono fattori importanti e ciò potrebbe in un alcuni specifici casi, prese le dovute accortezze, autorizzare alcune scelte di trattamento della scala a livello intervallare piuttosto che ordinale purché come dice Jamieson "questioni quali i livelli di misurazione e l'adeguatezza della media, della deviazione standard e delle statistiche parametriche dovrebbero essere considerate nella fase di progettazione e devono essere affrontate dagli autori"<sup>47</sup>.

Knapp dopo aver esaminato i contributi di vari suoi colleghi mettendoli insieme in un discorso ordinato termina invocando una tregua. La scienza però procede per dibattiti, dimostrazioni e confutazioni all'interno di un gruppo dove non sono ammesse tregue perché non servono come dice Kuhn "la conoscenza scientifica è intrinsecamente la proprietà comune di un gruppo o altrimenti non è assolutamente nulla. Per capirla dovremo conoscere le caratteristiche specifiche dei gruppi che la creano e la usano"<sup>48</sup>. Personalmente continuo ad essere dell'idea che la scienza non deve procedere per convenienza operativa quanto piuttosto per dimostrazioni e confutazioni. Nel momento in cui una misura non è quantificabile e si decide all'interno di una teoria rappresentazionale della misura di considerare come misurazione anche solo un'etichetta verbale non è logico snaturare tale processo data anche la non

---

<sup>46</sup> Knapp TR. *Treating ordinal scales as interval scales: an attempt to resolve the controversy*. Nurs Res 1990;123

<sup>47</sup> Jamieson S., *Likert scales: how to (ab) use them*. Medical Education 2004;38:1217–1218.

<sup>48</sup> Kuhn T. *The structure of scientific revolution*, International Encyclopedia of Unified Science, 1962, p251

dimostrabilità delle assunzioni che lo permetterebbero (come l'equidistanza degli intervalli).

Tra le varie problematiche che vi sono nel trattare la scala Likert come composta da variabili continue anziché discrete vi è anche la questione sulla validità. La validità di costrutto di uno strumento di misura indica quanto esso sia collegato al costrutto psicologico sottostante che si vuole misurare e in questo senso “l'analisi fattoriale gioca a ruolo centrale, in quanto è il metodo più applicato per ottenere evidenze sulla validità del costrutto perché fornisce informazioni sulla struttura interna dello strumento di misura”<sup>49</sup>. In uno studio simulativo Holgado e i suoi collaboratori hanno provato ad utilizzare alcune batterie di item di tipo Likert trattandole come fossero variabili continue, quindi a livello intervallo, sottoponendo poi il tutto ad analisi fattoriale. Essi concludono il loro lavoro spiegando che nonostante essi abbiano adottato tutte le precauzioni possibili perché ciò non si verificasse l'aver trattato tale scala a livello intervallo ha portato “a sua volta a problemi per quanto riguarda sia le correlazioni di Pearson che l'asimmetria nelle distribuzioni di risposte agli item”<sup>50</sup>.

### 3.2. Confutazioni

Lo scopo principale di questo elaborato è quello di mettere in luce come la violazione della significanza statistica nel modello Likert porti a delle problematiche non indifferenti di utilizzo della scala. Per fare questo abbiamo dovuto introdurre il tema della misurazione nelle scienze sociali per poi inserirci all'interno del dibattito sull'utilizzo di statistiche parametriche nel calcolo di variabili ordinali. È chiaro che in questa sede si sostiene la posizione per cui adottare modelli metrici per un sistema meramente classificatorio non sia opportuno ma in letteratura esistono opinioni differenti con cui è doveroso

---

<sup>49</sup> Zumbo, B. D. . *Validity: Foundational Issues and Statistical Methodology*. In C. R. Rao & S. Sinharay (eds.), (2007) *Handbook of statistics*, Vol. 26: Psychometrics (pp. 45-79). Amsterdam: Elsevier Science.

<sup>50</sup> Holgado-Tello, et al. *Confirmatory Factor Analysis of Ordinal Variables: A Simulation Study Comparing the Main Estimation Methods*, *Progressi nella psicologia latinoamericana* , (2018) 36 (3), 601

confrontarsi per poter avere una visione più ampia e possibilmente trovare una soluzione al problema.

Carifio e a Perla mettono in evidenza la differenza che sussiste tra l'item response set e la scala nel suo complesso criticando il fatto che molti ricercatori analizzano le problematiche sulla scala Likert attaccando di fatto la costruzione dei singoli item. Essi da questo punto di vista seguono un principio olistico per cui il tutto è più della somma delle parti quando dicono: "gli item della scala non sono autonomi e indipendenti (es. il comportamentista e il cieco empirista), ma piuttosto sono strutturati in un insieme ragionato"<sup>51</sup>. In particolare si appellano ad uno studio di Glass il quale "ha mostrato che l'F-test è estremamente robusto (salvo le violazioni dell'ipotesi di uguaglianza delle varianze) e che non si deve perdere potenza statistica utilizzando test statistici non parametrici quando al suo posto si possono analizzare i dati della scala Likert e persino analizzare tali dati selettivamente a livello di item"<sup>52</sup>.

Preso atto che non è dimostrabile l'equidistanza degli intervalli nelle categorie di risposta della scala Likert vi sono alcuni autori che sostengono che ciò non sia discriminante ai fini dell'accuratezza della scala e della precisione della misura. In particolare Geoff Norman riprendendo la tesi di Gaito e Lord per cui "i numeri non si ricordano da dove vengono"<sup>53</sup> (vedi capitolo 1.2) esprime chiaramente la sua posizione quando afferma che "anche se concettualmente una scala Likert è ordinale, nella misura in cui non possiamo teoricamente garantire che la distanza reale tra 1 = "Assolutamente in disaccordo" e 2 = "Non in disaccordo" è la stessa di "4 = "Nessuna opinione" e 5 = "Moderatamente d'accordo", questo è irrilevante per l'analisi perché il calcolo non ha modo di affermare o smentire ciò. Non ci sono osservazioni indipendenti per verificare o confutare il problema. E tutto ciò che il calcolo può fare è disegnare conclusioni sui numeri stessi. Quindi se i numeri sono distribuiti ragionevolmente possiamo fare inferenze"<sup>54</sup>. Egli conclude quindi il suo articolo affermando che "le

---

<sup>51</sup> Carifio J, Perla RJ. *Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes*. Journal of Social Sciences. 2007;3(3):p.109

<sup>52</sup> *ivi* p.110

<sup>53</sup> Lord, F.M. *On the statistical treatment of football numbers*. *American Psychologist*, 1953, p.750

<sup>54</sup> Norman, G., *Likert scales, levels of measurement and the "laws" of statistics* 2010 Springer Science+Business Media



statistiche parametriche possono essere utilizzate con dati Likert, con campioni di piccole dimensioni, con varianze non omogenee e con distribuzioni non normali, senza timore di arrivare a conclusioni errate"<sup>55</sup>

Altro punto focale della discussione è il fatto che i modelli metrici si basano su una distribuzione di probabilità normale o Gaussiana che i dati di matrice ordinale non possono garantire. Nonostante ciò Atkinson sostiene che "le analisi parametriche sono robuste come giudicato dall'osservazione che analisi parametriche e non parametriche portano a conclusioni simili in merito significato statistico"<sup>56</sup> anche se "la spiegazione di ciò potrebbe risiedere nel fatto che i numeri sono grandi e le distribuzioni sono simili"<sup>57</sup> perché a maggiore grandezza del campione nella porzione centrale le distribuzioni sono affini. "ANOVA e altri test di tendenza centrale sono altamente robusti a cose come l'asimmetria e la non normalità"<sup>58</sup>. Questi autori si basano sul principio della "robustezza" (robustness) della misura per cui si intende "il grado in cui il test darà la risposta giusta anche quando le assunzioni sono violate. E se non aumentano di molto le possibilità (o per niente), allora possiamo andare avanti"<sup>59</sup>. Labovitz per esempio spiega che "ci sono situazioni in cui è vantaggioso selezionare le tecniche classiche più potenti, anche se alcune ipotesi non sono soddisfatte, o la scala di misura non è esattamente intervallo o rapporto"<sup>60</sup>.

Si potrebbero riassumere le varie posizioni di questi autori nella visione per cui una "sopravvalutazione della critica fine a se stessa o dogmatismo statistico inappropriato"<sup>61</sup> limita le misurazioni nonostante sia dimostrata una certa robustezza dei dati.

---

<sup>55</sup> ibidem

<sup>56</sup>Mircioiu.C Atkinson.J *A Comparison of Parametric and Non-Parametric Methods Applied to Likert Scale* Pharmacy 2017 "p 10

<sup>57</sup> Ibidem

<sup>58</sup> Norman.G.,*Likert scales, levels of measurement and the "laws"of statistics* 2010 Springer Science+Business Media

<sup>59</sup> Ibidem

<sup>60</sup> Labovitz.S.*Some observations on Measurement and Statistics*,Oxford University Press,University of Southern California,1967,p10

<sup>61</sup> Bacchetti, P.. *Peer review of statistics in medical research: the other problem*. British MedicalJournal, 2002,234, 1271–1273.

### 3.3 Errori di primo e secondo tipo

Di fronte a tutte queste confutazioni sopra elencate alcuni autori si sono giustamente chiesti se sia veramente così innocuo trattare le variabili ordinali con modelli metrici. Uno di questi è Liddell che attraverso uno studio simulativo mette a confronto stessi campioni di dati trattandoli con due modelli statistici differenti al fine di confrontare eventuali differenze tra un'analisi dei dati svolta attraverso modelli metrici o un'elaborazione statistica eseguita attraverso l'utilizzo di ordered probit models per variabili ordinali. I risultati di questo studio in cui vengono riportati anche diversi casi di dati del mondo reale con caratteristiche molto simili ai dati simulati spiegano "perché ci sono discrepanze tra modelli metrici e modelli ordinali e illustrano come un'infinità di configurazioni creano falsi allarmi (errori di tipo I), mancanze (errori di tipo II) e inversioni di effetti"<sup>62</sup>.

Fino ad ora abbiamo osservato come l'impossibilità di dimostrare l'equidistanza tra le categorie di risposta della scala Likert porti ad una difficoltà nel considerare tale scala a livello intervallo rendendo pertanto prive di invarianza statistica tutte le operazioni eseguite a quel livello di scala in quanto i dati rimangono di matrice ordinale. Lo studio sopra esposto però aggiunge che nel trattare variabili ordinali secondo modelli metrici oltre a ostacoli relativi alla significanza statistica si sommano anche problemi sistematici relativi alla significatività statistica nel test di ipotesi. I ricercatori utilizzano tali test per verificare la validità delle loro ipotesi ed è una procedura di fondamentale importanza nel processo di costruzione della conoscenza. Il test di ipotesi seguendo il principio di falsificazione di Popper formula delle affermazioni che vengono considerate vere fintanto che non vengono smentite da ipotesi alternative. In questa verifica vi è la possibilità di commettere degli errori ovvero quello di rifiutare l'ipotesi nulla quando essa in realtà è vera oppure di accettarla quando essa è falsa. Nel primo caso si parla di errori del primo tipo o falsi

---

<sup>62</sup> Liddell et al(2018), *Analyzing ordinal data with metric models: What could possibly go wrong?* Department of Psychological and Brain Sciences,2018,p330

positivi mentre nel secondo caso di secondo tipo o falsi negativi. La significatività di un test d'ipotesi viene fissata arbitrariamente dal ricercatore e consiste nella probabilità di commettere errori del primo tipo quindi di rilevare cambiamenti tra due osservazioni che in realtà non esistono. Come sopra citato secondo Liddell l'adozione di modelli metrici per variabili ordinali crea sistematicamente falsi positivi, falsi negativi e inversione degli effetti nella formulazione del test di ipotesi inducendo quindi ad osservare variazioni che in realtà non ci sono o mancando nella rilevazione delle stesse.

Egli simula due gruppi composti da  $N=500$  nei quali vuole ipotizzare un'eventuale effect size tra due gruppi secondo il metodo di Cohen (1988) della differenza standardizzata tra medie (figura 13 in appendice). In figura 14 si può notare come l'effect size calcolato tramite l'ordered probit model tra i due gruppi sia pari a zero quindi c'è un'uguaglianza tra medie nei due gruppi mentre il modello metrico utilizzato fornisce una stima errata dei dati generando un evidente errore del primo tipo. Nella figura 15 invece si pone il caso di due gruppi le cui medie calcolate attraverso i modelli di calcolo ordinali sono sensibilmente differenti per cui di fatto esiste una differenza di effetto tra le due ipotesi la quale però non viene rilevata nel momento in cui si attuano statistiche parametriche.

Queste figure in appendice mostrano chiaramente come l'adozione di un modello metrico per variabili ordinali porti ad una dispersione dei dati tale che alcuni valori, soprattutto quelli più estremi, non vengono proprio rilevati. Infatti uno dei problemi principali dell'adozione di statistiche parametriche per variabili ordinali è il fatto che la relazione tra punteggi e misure è una relazione con andamento di tipo ogivale per cui i valori centrali sono lineari pertanto in tali valori punteggi e misure sembrano essere uguali. Tale dinamica però entra in crisi proprio nei valori estremi dove viene fuori la differenza tra l'andamento ogivale dei punteggi e quello lineare delle misure. Questo si vede molto bene nella figura 16 dove sono manifesti errori di primo, secondo tipo e inversione degli effetti nel test d'ipotesi.

## CONCLUSIONE

Quando si svolge una ricerca può spesso capitare di utilizzare determinati strumenti di rilevazione senza conoscerne realmente il funzionamento. Per quella che è stata la mia esperienza di studente fino ad ora mi sono accorto che tanti miei coetanei somministrano test perché lo devono fare senza interrogarsi realmente sull'efficacia di tali strumenti di misura. Per questa ragione sono contento di aver avuto la possibilità di poter approfondire il modello Likert secondo un principio di appropriatezza delle misure piuttosto che di robustezza o convenienza delle stesse.

Il dibattito sulla misurazione nelle scienze sociali continua ad essere ancora aperto nonostante sia passato più di un secolo dall'inizio di questa disputa ma è naturale se consideriamo che in matematica ci sono problemi irrisolti dai tempi di Euclide (congettura dei numeri primi gemelli). Il punto è che come abbiamo visto nel primo capitolo la psicologia si trova in un "vicolo cieco in cui ancora permane"<sup>63</sup> poiché "si limita a collegare numeri ad eventi solo sulla base di convenzioni formali come quella implicita nella famosa articolazione in livelli di scala proposta da Stevens"<sup>64</sup>. Ecco perché ci si appella alla necessità di adottare una misurazione fondamentale che sia in linea con i principi epistemologici della disciplina.

Il problema è che non solo "la definizione di Stevens ha permesso agli psicologi di presentarsi in un campo prestigioso come la scienza applicata"<sup>65</sup> ma anche che le regole che sono conseguenza di tale definizione vengono spesso violate. Abbiamo visto infatti come la violazione della significanza statistica nella scala Likert porti a dei problemi operativi non indifferenti. La mancata dimostrazione dell'equidistanza degli intervalli nelle categorie di risposta rende impossibile superare la condizione di ordinalità e anche se si

---

<sup>63</sup>Michell.J (1997,) *Quantitative science and the definition of measurement in Psychology*, British journal of Psychology cap 88 p. 361

<sup>64</sup> Michell J., (1999) *Measurement in psychology: a critical history of a methodological concept*, Cambridge University press

<sup>65</sup>Pitrone M.C (2012), *La misurazione fondamentale in psicologia in la sfida della misurazione nelle scienze sociali a cura di Paolo di Nicola*, Milano, FrancoAngeli s.r.l

applicasse ad ogni batteria di test l'analisi delle corrispondenze comunque resterebbe una scelta scorretta dal punto di vista metodologico. Vi sono peraltro alcuni limiti operativi evidenti che riguardano l'asimmetria nella distribuzione delle risposte agli item e il coefficiente di correlazione lineare di Pearson. Infatti qui non si tratta di ottenere una misura conveniente o precisa quanto piuttosto coerente con l'apparato epistemologico della scienza sottostante.

Se nell'ambito della statistica descrittiva questo aspetto è stato relativamente accettato, diverso è nel campo dell'inferenza dove in molti sostengono che "i numeri non si ricordano da dove vengono"<sup>66</sup> giustificando così l'adozione di statistiche parametriche nel calcolo di variabili ordinali. Nel terzo capitolo però è stato mostrato uno studio simulativo che ha messo in luce come l'adozione di modelli metrici per variabili ordinali porti sistematicamente ad errori di primo, secondo tipo ed inversione degli effetti nel test d'ipotesi.

La scala Likert è una delle più utilizzate nelle scienze sociali eppure anche una delle più criticate proprio per questa sua natura ibrida e semplificata. Ritengo però che il problema principale non sia tanto nella scala in sé quanto piuttosto nell'approccio alla misurazione. Come sopra espresso da Roberts(1987) il punto focale sta nell'appropriatezza della misura perchè media, deviazione standard e tutte le varie statistiche parametriche possono essere calcolate ovunque poi bisogna vedere quanto sia appropriato fare certe affermazioni su tali dati.

---

<sup>66</sup> Lord, F.M. *On the statistical treatment of football numbers*. *American Psychologist*, 1953, p.750

## APPENDICE

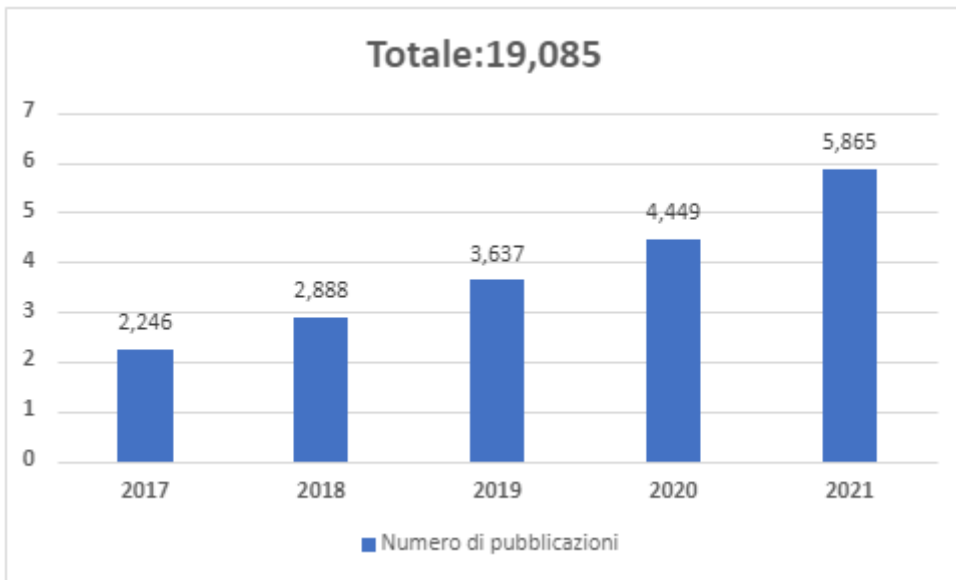


Tabella 1: Grafico che mostra il numero di articoli degli ultimi cinque anni aventi il termine Likert come parola chiave (fonte: Scopus). Si può assistere ad un utilizzo sempre crescente di tale strumento di misura soprattutto nell'ultimo anno.

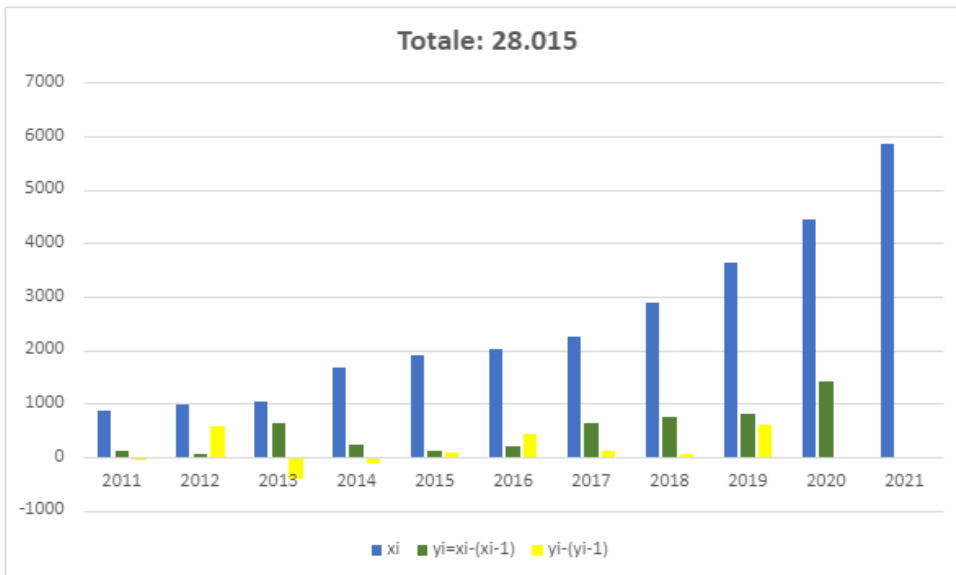


Tabella 2: Stessa tipologia del grafico in tabella 1 dove:

$x_i$  = numero pubblicazioni

$y_i = x_i - (x_{i-1})$  = differenza del numero degli articoli pubblicati tra l'anno successivo e quello precedente

$y_i - (y_{i-1})$  = variazione delle differenze del numero degli articoli pubblicati tra l'anno successivo e quello precedente

Scale	Basic Empirical Operations	Mathematical Group Structure	Permissible Statistics (invariantive)
<b>NOMINAL</b>	Determination of equality	<i>Permutation group</i> $x' = f(x)$ $f(x)$ means any one-to-one substitution	Number of cases Mode Contingency correlation
<b>ORDINAL</b>	Determination of greater or less	<i>Isotonic group</i> $x' = f(x)$ $f(x)$ means any monotonic increasing function	Median Percentiles
<b>INTERVAL</b>	Determination of equality of intervals or differences	<i>General linear group</i> $x' = ax + b$	Mean Standard deviation Rank-order correlation Product-moment correlation
<b>RATIO</b>	Determination of equality of ratios	<i>Similarity group</i> $x' = ax$	Coefficient of variation

Tabella 3: Stevens indica la definizione, la struttura matematica e le statistiche significative per ogni tipologia di scala da lui proposta

	AMBITO	POPOLAZIONE	N	FIGURA
FUMO (Computel, 1987) atteggiamento verso il fumo	Italia	adulti	4994	8
SURITA680 (Barnes, Sani, 1968) atteggiamento verso la politica e le istituzioni	Italia	adulti	1925	12
SURITA72 (Barnes, Sani, 1972) atteggiamento verso la politica e le istituzioni	Italia	adulti	1841	9
CIECHI (Doxa, 1989) atteggiamento verso i non vedenti	Italia	adulti	1085	4
INSEG (Iard, 1990) opinioni sulla scuola italiana	Italia	insegnanti	5000	5
RELIG (Eurisko, 1988) atteggiamento verso la religione	Italia	giovani 19-25 anni	1473	2
FONDI (Computel, 1990) opinioni sui fondi comuni d'investimento	Italia	adulti 25-60 anni	504	11
SURGIOLO (Tullio-Altan, Cartocci, 1976) valori dei giovani lombardi	Lombardia	giovani 14-25	801	3
EMILIA (Sapignoli, Seligardi, 1988) atteggiamento verso emancip. femminile, immigrazione meridionale e sindacati	Emilia-Romagna	adulti emiliani	200	7
TOSCANA (Giovani, Pellizzari, 1989) atteggiamento verso l'ambiente e l'emancipaz. femminile	Toscana	adulti metà laureati e metà con la licenza media	100	10
VERONA (Morlandi, Secondulfo, 2003) valori degli universitari	Verona	studenti universitari	444	6

Tabella 4: I questionari analizzati da Macrì i quali verranno successivamente sottoposti all'ac

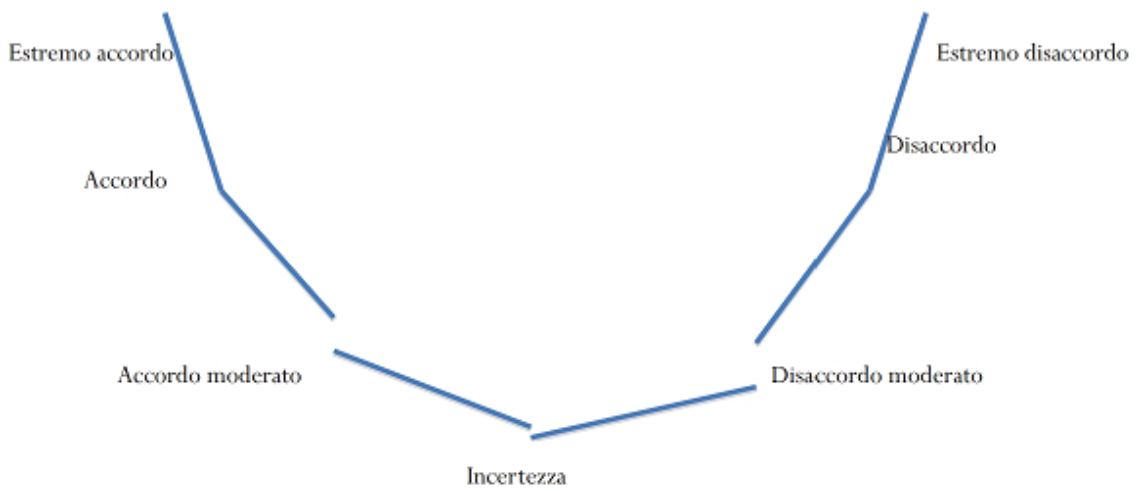


Figura 1 : *Ipotetico diagramma prodotto dall'ac se le categorie di risposta alle scale fossero equidistanti*

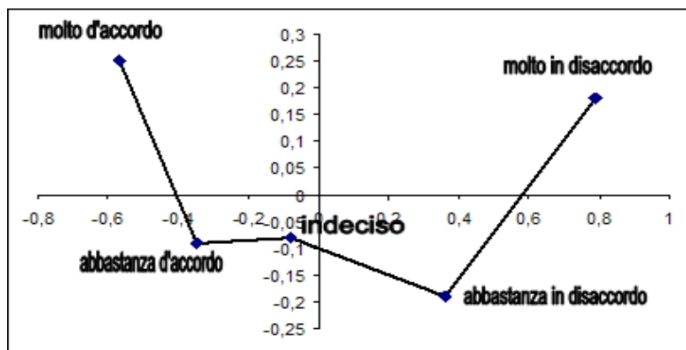


Figura 2: *ac applicata ad un questionario sull'atteggiamento verso la religione (1473 casi)*

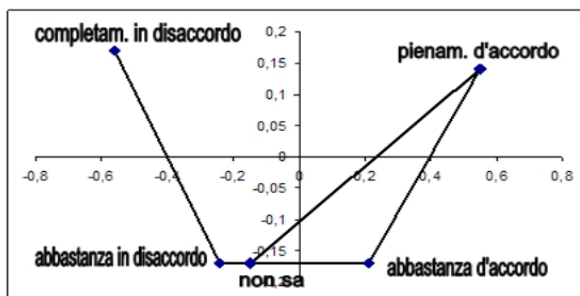


Figura 3: *Risultati dell'ac al questionario sui Valori dei giovani in Lombardia (801 casi)*



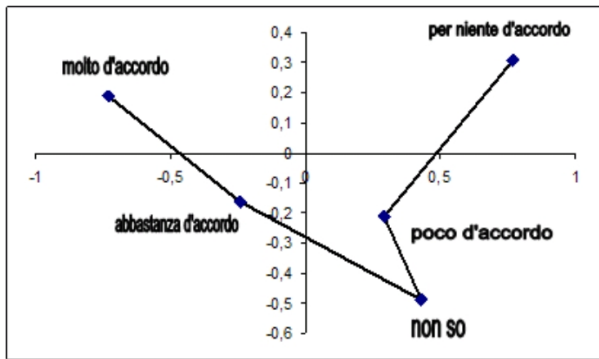


Figura 4: *ac* applicata ad un questionario sull' atteggiamento verso i non vedenti (1085 casi)

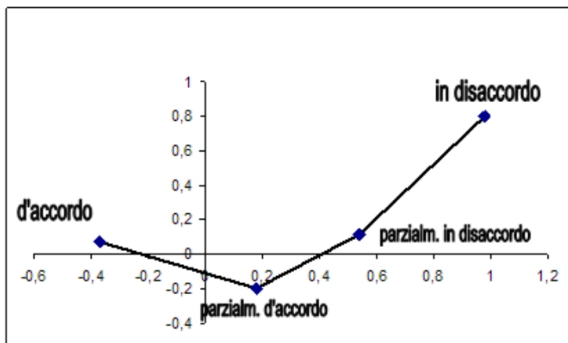


Figura 5: Questionario IARD sulle opinioni degli insegnanti sulla scuola italiana(5.000 casi)

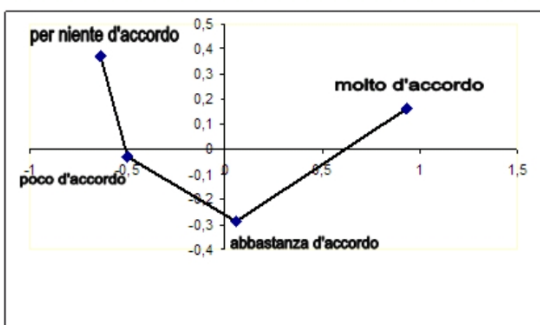


Figura 6: *ac* applicata ad un questionario sui Valori degli studenti universitari di Verona (444 casi)

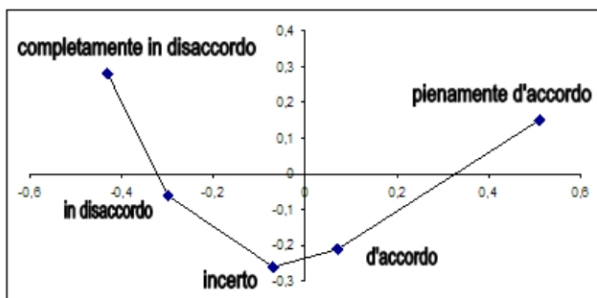


Figura 7: Risultati dell'*ac* al questionario sull'emancipazione femminile,immigrazione meridionale e ruolo dei sindacati.(200 casi)

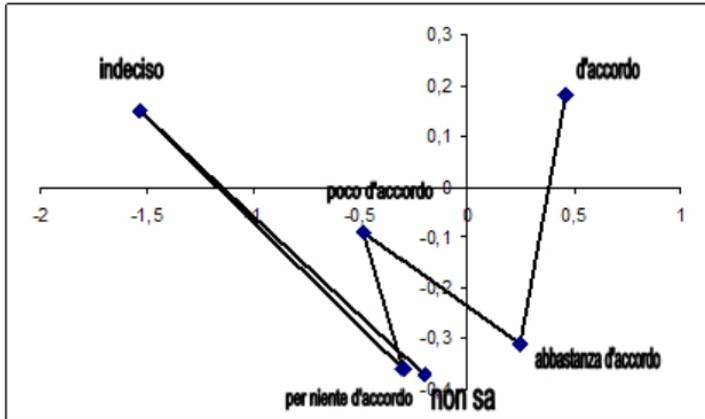


Figura 8: Risultati dell'ac applicati ad un questionario sull'atteggiamento verso il fumo(4994 casi)

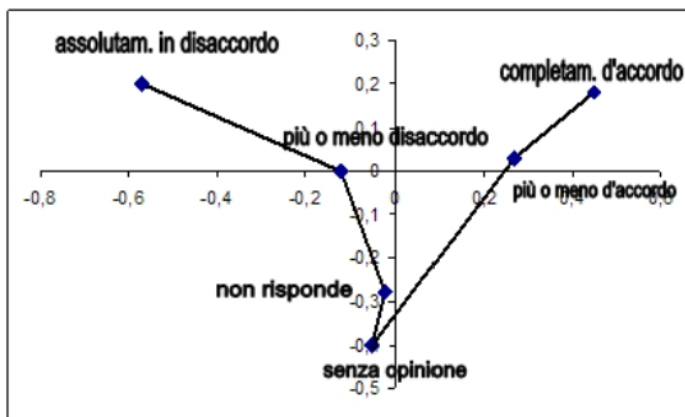


Figura 9: ac applicata ad un questionario sull'atteggiamento verso la politica e le istituzioni(1841 casi)

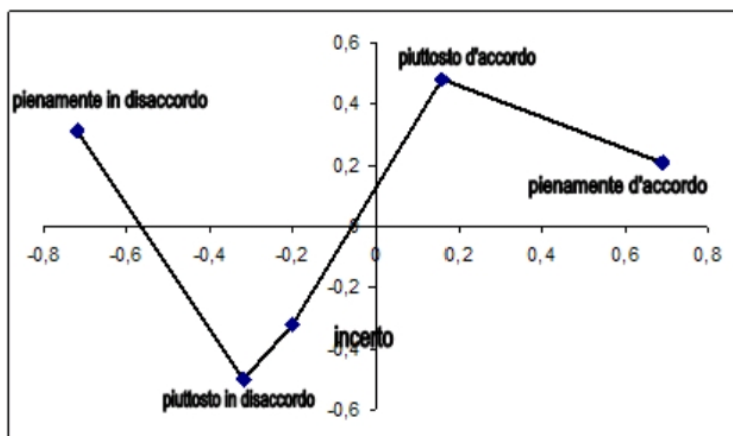


Figura 10: Questionario sull'ambiente e l'emancipazione femminile.(100 casi)

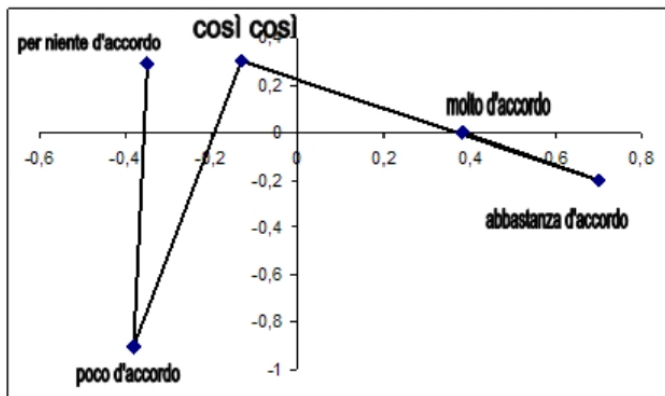


Figura 11: Risultati dell'ac applicata ad un questionario sui fondi di investimento(504 casi). Da notare la posizione anomala delle categorie così così ed abbastanza d'accordo che risultano essere più impegnative ed estreme del previsto non rispettando il presunto ordine delle categorie di risposta

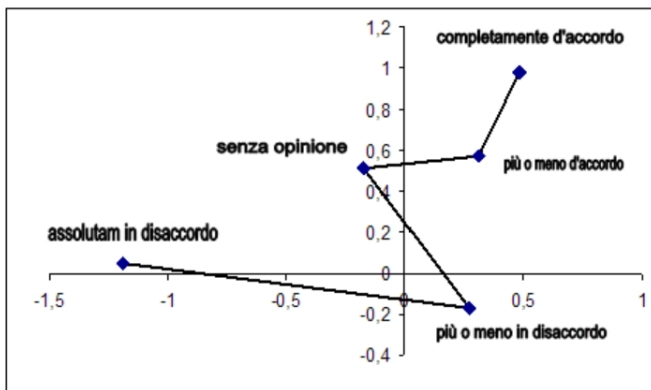


Figura 12: ac su di un questionario sull' atteggiamento verso la politica e le istituzioni (1925 casi). Anche in questo caso l'ordine delle categorie è stravolto e se confrontato con la figura 9(stesso questionario campione diverso) risulta altrettanto insolita la posizione della categoria senza opinione.

$d = (\mu_1 - \mu_2) / \sqrt{(\sigma_1^2 + \sigma_2^2)/2}$ . Figura 13: Differenza standardizzata tra medie di Cohen (1988)

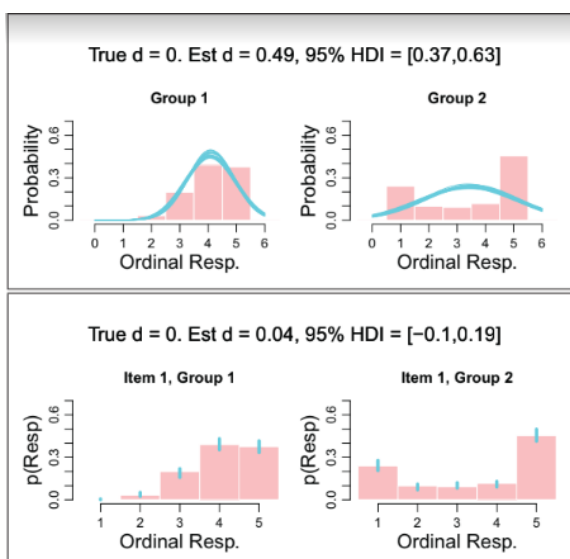


Figura 14: Liddell confronta due gruppi tramite il test di verifica di ipotesi adottando sia il modello metrico che quello per variabili ordinali. Si può notare un evidente rifiuto dell'ipotesi nulla quando essa è vera nel momento in cui viene adottato il modello metrico con conseguente errore del primo tipo.

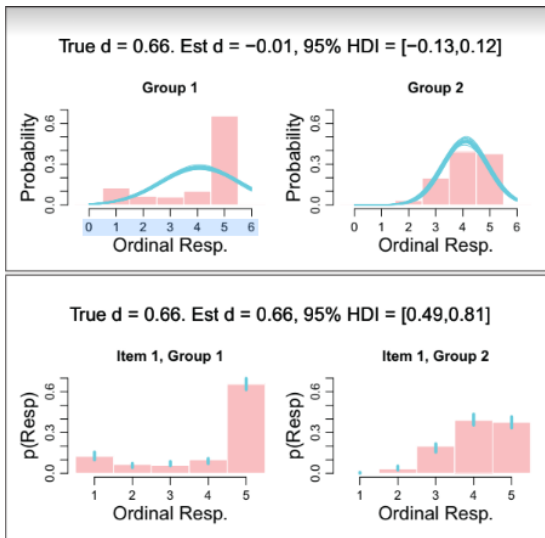


Figura 15: Sempre nello stesso studio della precedente figura invece troviamo un manifesto errore di secondo tipo con mancanza di rilevazione dei dati soprattutto nei valori estremi. Il modello metrico dimostra di essere meno adatto a questo test di verifica.

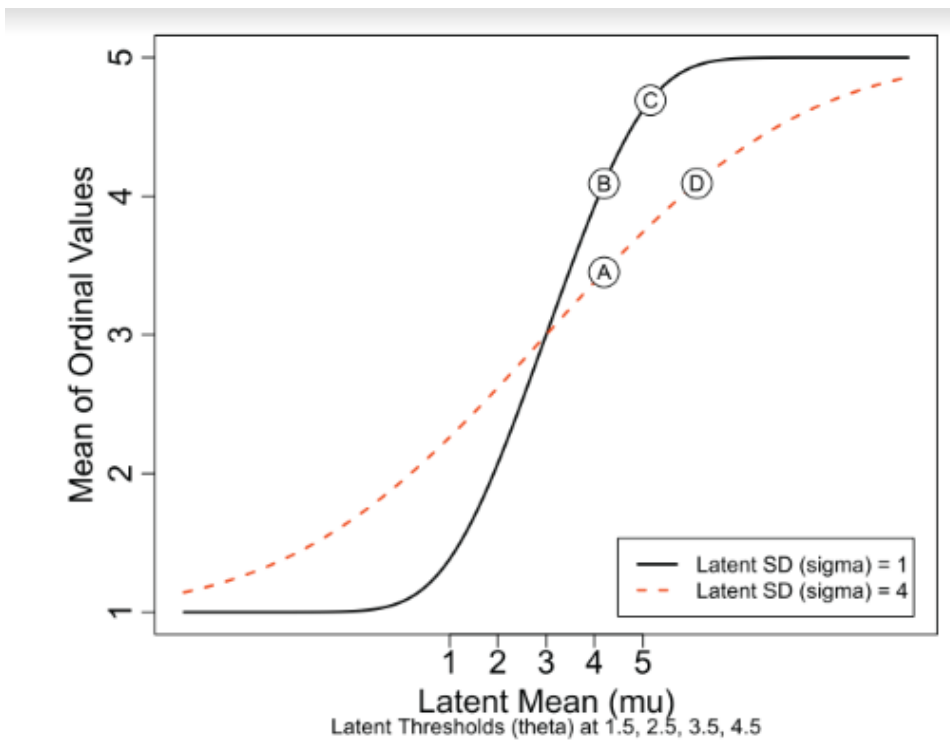


Figura 16: Grafico che mostra le medie di variabili ordinali come funzione della media latente ( $\mu$ ). I gruppi A e B hanno la stessa media latente ma diversa media ordinale (errore di primo tipo). I gruppi B e D invece hanno la stessa media ordinale ma diversa latenza (falsi positivi). C e D invece mostrano inversione degli effetti.

## BIBLIOGRAFIA

Alwin, . (1997). *Feeling thermometers versus 7-point scales*. Sociological Methods and Research

Amaturo E. (1989), *Analyse des données e analisi dei dati nelle scienze sociali*, Torino: Centro Scientifico.

Bacchetti, P. (2002). Peer review of statistics in medical research: the other problem. British Medical Journal

Baggaley.A,Hull.A(1983),*The effect of non linear transformations on a Likert scale*,Evaluation of health professions

Benzécri J. P. (1973; II ed. 1982), *L'analyse des données*, vol. II, L'analyse des correspondances, Paris: Dunod.

Bruschi A., (1999), *Metodologia delle scienze sociali*, Milano, Bruno Mondadori.

Carifio J, Perla RJ.(2007) *Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes*. Journal of Social Sciences

Colman AM, et al(1997) *Comparing rating scales of different lengths: Equivalence of scores from 5-point and 7-point scales*. Psychological Reports.;

Coombs, Clyde H. [1950] *Psychological Scaling Without a Unit of Measurement*, in "Psychological Review", vol. LVII, n. 3. - [1964] *A Theory of Data*, New York, Wiley

Di Franco G., Marradi A. (2003), *Analisi dei fattori e analisi in componenti principali*, Acireale-Roma: Bonanno.

Ferguson et al., (1940), *Quantitative estimates of sensory events: final report*, Advancement of science.

Fuso.S(1999), *Realtà o illusione? Scienza, pseudoscienza e paranormale*, Edizioni Dedalo,.

Gaito,J.(1980). *Measurement scales and statistics: Resurgence of an old misconception*, Psychological Bulletin.

Glass, G.V., P.D. Peckham and J.R. Sanders(1972), *Consequences of failure to meet assumptions underlying the analyses of variance and covariance*. Review of Educational Research,

Hodge, D. R., & Gillespie, D. F. (2007). *Phrase completion scales: A better measurement approach than Likert Scales?*.Journal of Social Service Research,

Holder O. L., (1901), Die Axiome der Quantität und die Lehre vom Mass, *Berichten der mathematisch-physischen Classe der Königl. Sächs. Gesellschaft der Wissenschaften zu Leipzig*,, Eng. trans. J. Michell and C. Ernst, "The Axioms of Quantity and the Theory of Measurement", *Journal of Mathematical Psychology*,.

Jordan.N (1965), *The Asimmetry of "Liking" and "Disliking": A Phenomenon Meriting Further Reflection and Research* Public Opinion Quarterly,

Jamieson S(2004). Likert scales: how to (ab) use them. Medical Education ;38:1217–1218.

Joshi A. et al.(2015) *Likert scale:Explored and explained* .British journal of applied science &technology

Holgado-Tello, FP, Morata-Ramirez, M. Angeles e Barbero Garcia, MI (2018). *Confirmatory Factor Analysis of Ordinal Variables: A Simulation Study Comparing the Main Estimation Methods*,*Progressi nella psicologia latinoamericana* , 36 (3), 601-617

Knapp TR(1990). *Treating ordinal scales as interval scales: an attempt to resolve the controversy*. Nurs Res

Kuzon WM et al(1996) *The seven deadly sins of statistical analysis*. Ann plast surg

Labovitz.S.(1967)*Some observations on Measurement and Statistics*,Oxford University Press,University of Southern California,

Lalla.M.(2015),*Le scale ordinali e i relativi problemi operativi*,Archivio istituzionale della ricerca - Università di Modena e Reggio Emilia

Larroulet Philippi, C. (2021). On Measurement Scales: Neither Ordinal Nor Interval? *Philosophy of Science*

Liddell et al(2018),, *Analyzing ordinal data with metric models: What could possibly go wrong?* Department of Psychological and Brain Sciences

Likert, R(1932). *A Technique for the Measurement of Attitudes*. Archives of Psychology,

Luce R.D.,TurkeyJ.W(1964), *Simultaneous conjoint measurement:a new type of fundamental measurement*,Journal of mathematical Psychology

Lord,F.M.(1953) *On the statistical treatment of football numbers*,.American Psycholgy

Marradi A. Macrì.E(2012) *Sono equidistanti le categorie di una scala Likert?*  
Cambio:rivista sulle trasformazioni sociali

Marradi A., Gasperoni G. ,(2002) *Costruire il dato 3. Le scale Likert*, Milano Franco Angeli

Marradi A. Macrì.E (2012) *Sono equidistanti le categorie di una scala Likert?*  
Cambio:rivista sulle trasformazioni sociali

Michell.J (1997,) *Quantitative science and the definition of measurement in Psychology*,Brititsh journal of Psychology

Michell J., (1999) *Measurement in psychology:a critical history of a methodological concept*,Cambridge University press

Michell J., (2007) *Representational theory of measurement*. Amsterdam: Elsevier Science

Mircioiu.C Atkinson.J(2017) *A Comparison of Parametric and Non-Parametric Methods Applied to Likert Scale Pharmacy*

Naia A. et al(2022) *How rating scales influence responses' reliability, extreme points, middle point and respondent's preferences*, *Journal of Business Research*

Norman.G(2010)., *Likert scales, levels of measurement and the "laws"of statistics*  
Springer Science+Business Media

Ogden J, Lo J(2012)*How meaningful are data from Likert scales? An evaluation of how ratings are made and the role of the response shift in the socially disadvantaged. Journal of Health Psychology.*

Park K.(2013) *Social sciences and health*. In Park K, editor. Park Text Book of Preventive and Social Medicine. Jabalpur: Bhanot Publishers.

Pell.G.(,2005,) *Use and misuse of Likert scales*, Medical Education

Ramsay J.O.(1975), *Review of foundation of measurement*, Vol I, by Krantz D.H et al  
Psychometrika

Reese T.W. (1943) *The application of the theory of physical measurement to the measurement of psychological magnitudes, with three experimental examples*, Psychological Monographs

Regenwetter, M., Hsu, Y.-F., & Kuklinski, J. H. (2019). *Towards meaningful inferences from attitudinal thermometer ratings. Decision*

Roberts M. et al.(1987) *Meaningless Statistics* Journal of Education statistics

Robusto.E.(2012) *La misurazione fondamentale in psicologia in la sfida della misurazione nelle scienze sociali a cura di Paolo di Nicola*, Milano, FrancoAngeli s.r.l

Stevens S. S., (1946), *On the theory of scales of measurement*, Science.

Townsend J. T. and Ashby F. G., (1984), *Measurement scales and statistics: The misconception misconceived*. Psychological Bulletin.

Weinbach, Robert W. (1989) *"When Is Statistical Significance Meaningful? A Practice Perspective," The Journal of Sociology & Social Welfare*

Wilcox.C. et al(1989), *Some like it hot individual differences in responses to group feeling thermometers*, *Public Opinion Quarterly*

Wu,S.Leung,(2017) *Can Likert Scales be treated as Interval Scales-A simulation study*, Journal of Social Service Research

Zumbo, B. D. (2007). *Validity: Foundational Issues and Statistical Methodology*. In C. R. Rao & S. Sinharay (eds.), *Handbook of statistics*,