

Indice

| | |
|------------------------------------------------------------------------------|-------|
| I. Introduzione | p. 3 |
| I.1 Premessa..... | p. 3 |
| I.2 Motivazioni ed obiettivi..... | p. 3 |
| I.3 Descrizione del caso di studio..... | p. 4 |
| I.4 Metodi e tecniche | p. 4 |
| I.5 Organizzazione della tesi..... | p. 6 |
| 1. <i>Discrimination background</i> | p. 7 |
| 1.1 La discriminazione..... | p. 7 |
| 1.1.1 I gradi della discriminazione e i gruppi protetti..... | p. 8 |
| 1.1.2 La discriminazione diretta e indiretta..... | p. 9 |
| 1.1.3 Le cause della discriminazione..... | p. 10 |
| 1.1.4 Le azioni affermative, la discriminazione inversa e i favoritismi..... | p. 11 |
| 1.2 La discriminazione in ambito legale..... | p. 12 |
| 1.2.1 Le prove statistiche dei casi discriminatori..... | p. 13 |
| 1.2.2. La legislazione sulla discriminazione..... | p. 14 |
| 1.3 La discriminazione in ambito economico..... | p. 19 |
| 1.3.1 I modelli economici..... | p. 20 |
| 1.4 I pregiudizi di genere nella ricerca scientifica..... | p. 22 |
| 2. La base di dati..... | p. 24 |
| 2.1 Il <i>database</i> del FIRB..... | p. 24 |
| 2.2 Le basi di dati..... | p. 24 |
| 2.2.1 La descrizione delle basi di dati..... | p. 25 |
| 2.3 Le modalità di finanziamento del progetto..... | p. 26 |
| 3. <i>Discrimination Discovery</i> : approccio con l'analisi dei dati..... | p. 32 |
| 3.1 Analisi dei dati per analisi discriminatorie..... | p. 32 |
| 3.2 Analisi preliminari..... | p. 32 |
| 3.2.1 Analisi dei costi progetto..... | p. 35 |
| 3.3 <i>Discrimination Discovery</i> | p. 42 |
| 3.3.1 Analisi dell' <i>International peer-review</i> | p. 43 |
| 3.3.2 Analisi dei costi e produttività <i>step 1</i> | p. 49 |

| | | |
|-------|------------------------------------------------------|-------|
| 3.3.3 | Distribuzione dei voti..... | p. 50 |
| 3.3.4 | Analisi dell' <i>Audit</i> | p. 63 |
| 3.3.5 | Analisi dei costi e produttività <i>step 2</i> | p. 69 |
| 3.3.6 | Analisi di regressione..... | p. 72 |
| 4. | Conclusioni..... | p. 94 |
| | Bibliografia..... | p. 97 |

Introduzione

I.1 Premessa

La donna nella storia della civiltà è sempre stata subordinata all'uomo. Le differenze tra i due sessi hanno portato il sesso maschile a prevalere su quello femminile, occupando un ruolo privilegiato all'interno della società. La donna fino dai tempi antichi è sempre stata posta in una condizione di inferiorità, evolvendosi in una società essenzialmente misogina. Molti pregiudizi, che vivono ancora oggi nell'immaginario collettivo, hanno radici molto lontane e sono stati influenzati dal pensiero di molti filosofi, letterati e politici.

Nel XX secolo lotte femministe per la parità dei sessi crebbero ad un'alta velocità, ma nonostante ciò tutt'oggi le donne continuano ad essere molto sottovalutate e sfruttate. La condizione della donna nel mondo è sempre stata caratterizzata da una certa inferiorità sia sul piano sociale, che su quello giuridico, politico e professionale, anche se ai nostri giorni la differenza tra il sesso maschile e quello femminile è sicuramente molto meno evidente soprattutto nel mondo occidentale.

Tema caldo dell'attualità è la discriminazione sul mondo del lavoro tra i due sessi, sia in termini di remunerazione, sia di opportunità lavorative. In questo lavoro si vedrà in particolare la discrepanza per quanto riguarda gli incentivi sui progetti di ricerca

I.2 Motivazione ed obiettivi

La motivazione principale di questo lavoro è capire quanto sono discriminate le donne negli ambiti di ricerca, per quanto riguarda i contributi su ogni progetto da esse diretti, rispetto ai progetti diretti da uomini; quindi il filo rosso di questa ricerca è comprendere se le donne hanno le solite possibilità di ricevere fondi, esattamente come gli uomini. Conseguentemente l'obiettivo è quello di ispezionare i dati del FIRB (Fondo per gli Investimenti della Ricerca di Base), relativi alle richieste di finanziamento sul programma FIRB giovani 2008 in base al sesso dei coordinatori. Detto questo è opportuno dire che si considererà il sesso dei relativi coordinatori come la risorsa principale per investigare la discriminazione dei due sessi.

I.3 Descrizione del caso di studio

Il punto di partenza per delineare il caso di studio è consistito nella comprensione della definizione giuridica di discriminazione e degli approcci che le varie legislazioni intraprendono per regolamentarla. È stata studiata, per questo fine, la legislazione dell'Italia che prende spunto dalla Dichiarazione Universale dei Diritti dell'Uomo e dalla Carta dei Diritti Fondamentali dell'Unione Europea. Inoltre non è stato semplice trovare una definizione di discriminazione a livello sociale che ne comprendesse tutte le forme ed i contesti, nonostante il problema sia già diventato rilevante da sempre, in particolare nel caso delle donne.

Per comprendere a pieno le difficoltà del problema in questione si illustreranno brevemente i processi, nonché perizie, che devono affrontare i progetti prima di ricevere un eventuale finanziamento. Bisogna precisare che ogni unità di ricerca ha un coordinatore che dirige il progetto; i direttori si distinguono in due principali unità di ricerca: giovani e strutturati. L'accettazione finale di un progetto è sottoposto a due principali *step*: il primo è *international peer-review* e in seguito *audit*. Se entrambi le due fasi hanno riscontro positivo il MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca) classifica l'insieme dei coordinatori e decide sia se il progetto sarà avviato, sia l'entità dei fondi concessi. Quindi dato che un progetto segue una prassi prima di essere finanziato, la complessità del caso di studio sta nel capire dove e quando inizia la vera e propria discriminazione. Perciò prima si dovrà valutare la prevenzione sulla discriminazione e successivamente valutare la presenza di discriminazione. Questi concetti saranno delucidati nella sezione successiva dove si introdurranno i metodi e le tecniche con le quali si è affrontato il problema.

I.4 Metodi e tecniche

Il fenomeno sociale della discriminazione è stato ampiamente investigato nel passato secondo diverse prospettive: economiche, legali e sociologiche. Recentemente tale campo è diventato un settore d'interessante scoperta non solo per le materie classiche, ma anche per nuove discipline come la *computer science*.

Negli ultimi decenni si sono creati strumenti utili alla memorizzazione di enormi quantità di dati; è nata quindi la necessità di sviluppare nuove tecnologie

indispensabili per l'analisi dei dati e l'elaborazione di questi dati che altrimenti non avrebbero avuto nessuna utilità concreta. Il *data warehousing* e il *data mining* nascono in questa prospettiva. Il primo è una collezione di metodi, tecnologie e strumenti di ausilio al cosiddetto "*knowledge worker*" per condurre analisi di dati finalizzate all'attuazione di processi decisionali e al miglioramento del patrimonio informativo. Nello specifico si analizzeranno i dati secondo la fruizione OLAP (*on-line analytical processing*) che consiste nell'analisi multidimensionale dei dati tramite un percorso di navigazione a più dimensioni, secondo le operazioni tipiche di *roll-up*, *roll-down*, *slice and dice*. Il secondo strumento di analisi è il *data mining*, utile per scoprire informazioni nascoste nei dati. Gli scopi per cui vengono utilizzate le tecniche di *data mining* sono ormai diversi e variano dalla semplice analisi statistica, allo studio di settore, sino al supporto alle decisioni. Il *data mining* fa parte di un più ampio processo chiamato *Knowledge Discovery in Database* (KDD) che, come si evince dal nome, ha lo scopo di estrarre la conoscenza dai dati. È chiaro che il dato è il risultato di una misurazione di un certo evento e di per se non ha una grande utilità. Tuttavia valutando i dati in un preciso contesto e applicandovi opportune elaborazioni, è possibile ricavare informazioni: in questo modo il dato diventa utile e in grado di rispondere ad una richiesta specifica.

La fase di *Data pre-processing* verso il database del FIRB è stata quella di applicare un *data cleaning* che comporta una aumento della qualità dei dati tramite: aggiunta di valori mancanti, aggiustamento dei dati rumorosi, identificazione ed eliminazione di *outliers* e soluzione delle inconsistenze, perché senza dati di qualità non è possibile effettuare un'analisi di qualità. Successivamente si è passati alla fase delle analisi preliminari. In questo processo si descrivono le proprietà dei due *data mart* tramite le analisi delle sue variabili più caratterizzanti e le analisi dei costi.

In seguito si è passati alla fase di *discrimination discovery*. In questa fase è stata svolta una verifica sulla qualità delle attività di ricerca in base al risultato ottenuto nella fase di *international peer-review* e *audit* e in base ad alcune variabili dei costi progetto, per verificare se i progetti non ammessi fossero effettivamente non idonei. Per verificare questa ipotesi si è innanzitutto controllato i costi per valutare il costo medio dei progetti idonei e non, per poi valutare lo *status* dei progetti secondo la produttività. In secondo luogo si è analizzato come varia la frequenza

dell'*international peer-review* rispetto al crescere/decrescere del valore stesso, utilizzando i metodi della statistica descrittiva. Inoltre tramite misure di associazione, è stata valutata che tipo di relazione vi fosse tra variabili nel *database* del FIRB. Questa fase di *discrimination discovery* è stata svolta per mezzo di operazioni OLAP sulle unità *audit e peer-review*, svelando una possibile presenza di discriminazione.

Quindi riepilogando, l'analisi per svelare la discriminazione ai danni di capi progetto donne è stata svolta analizzando interamente il processo di selezione per l'erogazione fondi sulle attività di ricerca e sui costi. Sul processo di selezione progetti per i fondi è stata applicata la *discrimination discovery* per scovare una possibile discriminazione.

I.5 Organizzazione della tesi

Questo lavoro parte dalla definizione giuridica di discriminazione per poi passare a quella economica e, attraversando varie fasi, arriva a definire un ambiente di analisi basato sull'analisi dei dati.

- Capitolo 1: viene presentata una rassegna delle fonti legislative / giurisprudenziali ed economiche, per poi passare al compendio di alcuni lavori economico-statistici che faranno luce sul fenomeno della discriminazione nei sessi.
- Capitolo 2: si descrive la struttura della base di dati relativa al database del FIRB-10. Bisogna precisare che le due basi di dati su cui vengono effettuate le analisi sono un *data-mart* che sintetizzano il database del FIRB.
- Capitolo 3: si spiega lo stato dell'arte e successivamente il modello sviluppato per il trattamento e l'analisi delle variabili nel database FIRB per un lavoro di *discrimination discovery*. Inoltre vuole essere una rassegna dei risultati ottenuti dai processi di analisi di dati sul database FIRB.
- Capitolo 4: si illustreranno le conclusioni alle quali si è arrivati dopo aver analizzato i dati del *database* FIRB relativi alla discriminazione tra i due sessi.

1. *Discrimination background*

1.1 La discriminazione

Il termine discriminazione si riferisce ad un ingiustificato trattamento non paritario nei confronti di individui facenti parti di uno stesso gruppo o categoria¹. Le distinzioni giustificate sono quelle permesse dalla legge. Il termine stesso discriminazione, in quanto atto del discriminare, ovvero “distinguere una o più cose o persone da altre, far differenza”², non ha etimologicamente una connotazione negativa, anche se nel linguaggio comune ormai ha questa interpretazione; tuttavia la assume nel momento in cui si ricorre a caratteristiche superficiali per ridurre la possibilità degli individui di accedere alle stesse opportunità economiche, sociali e politiche. Per superficialità delle caratteristiche si intende che queste siano parzialmente o completamente slegate dalle abilità, dai talenti e dalle forze potenziali o effettive degli individui stessi. Quindi in una società si ha discriminazione quando le compensazioni sono almeno in parte distribuite sulla base di caratteristiche estranee alla reale produttività dei soggetti.

L'atto del discriminare sempre è legato ad un pregiudizio (dal latino *prae*, "prima" e *iudicium*, "giudizio"³), ossia un giudizio prematuro basato su argomenti insufficienti o su una loro non completa o indiretta conoscenza. Di primaria importanza si è rivelato il problema di accertare la presenza, la misura, la natura e il grado di discriminazione nell'attuare decisioni politiche economiche o sociali con il fine ultimo di prevenirla. Infatti negli ultimi cinquanta anni il fenomeno della discriminazione è stato ampiamente investigato secondo diverse prospettive: sociali, legali, economiche; recentemente tale campo è diventato un settore di interessante scoperta non solo secondo prospettive classiche ma anche per nuove discipline, come la scienza computazionale. Infatti oggi per studiare il tema della discriminazione ci si è avvalsi anche della scienza computazionale, orientata all'applicazione e alla

1 Romei e Ruggieri, *A multidisciplinary survey on discrimination data analysis*, p. 2.

2 Zingarelli, *lo Zingarelli 2004*, *vocabolario della lingua italiana*, p. 558.

3 <http://it.wikipedia.org/wiki/Pregiudizio>

soluzione di problemi basandosi molto sull'impiego di calcoli computazionali al fine di risolvere problemi inaccessibili per i tempi e le modalità di calcolo umani; infatti ciò è stato possibile grazie alla comodità di incamerare e memorizzare grandi quantità di dati, al continuo incremento della potenza dei computer, allo sviluppo del *data analysis* e delle tecniche di *mining* con il fine di migliorare tecniche statistiche classiche ed econometriche.

1.1.1 I gradi della discriminazione e i gruppi protetti

Le leggi sui diritti umani proibiscono la discriminazione sulla base del genere, etnia, razza, orientamento sessuale, colore della pelle, origine sociale, lingua, religione, opinioni politiche o altre opinioni personali, caratteristiche genetiche, parentela, disabilità, malattia, età e stato civile come vedremo in seguito nella legislazione italiana. Le legislazioni nazionali e internazionali rivolgono attenzioni diverse a particolari categorie di gruppi, molto spesso riservate solo a specifiche caratteristiche: è questo il caso di assicurazioni, sconti, affitti, prestiti e mutui, concorsi e bandi pubblici e persino nell'ambito dell'educazione. A livello generale comunque la discriminazione si concentra su qualunque trattamento culturale o fisico non paritario proprio nella vita di tutti i giorni.

Con il termine gruppo protetto ci si riferisce a un particolare gruppo o categoria di persone, prese in considerazione come soggetto dell'analisi discriminatoria, indipendentemente dal fatto che nessuna legge nazionale o internazionale esplicitamente protegge un gruppo dalla discriminazione in ambiti specifici presi in considerazione⁴.

Quando parliamo di discriminazione molto spesso si tende a pensare alle varie forme in cui si presenta: verso le donne, gli stranieri, i disabili, e così via. Tuttavia, anche se i gradi della discriminazione vengono di solito considerati separatamente (genere, razza, disabilità), l'interazione di diverse forme di discriminazione sta diventando un ambito molto interessante e quindi studiato. Ad esempio una donna anziana disabile può essere soggetta ad un trattamento discriminatorio nei confronti di una persona più giovane, di un'altra donna, di un'altra disabile o nei confronti di più categorie combinate tra loro. Si definisce un atto di *discriminazione multipla*

⁴ Romei e Ruggieri, *A multidisciplinary survey on discrimination data analysis*, p. 2.

quando una persona è discriminata sulla base di diverse caratteristiche considerate nello stesso momento: così anche se ogni singola discriminazione agisce indipendentemente secondo esperienze distinte, la discriminazione multipla si riferisce all'impatto cumulativo delle singole discriminazioni; ossia si è discriminati per più di un motivo. Si parla invece di *discriminazione composta* quando differenti gradi di discriminazione sono considerati nello stesso momento; per esempio nel caso di donne immigrate che sono maggiormente disoccupate (come immigrate rispetto ai residenti) e sottopagate (come le donne lavoratrici rispetto agli uomini).

Molte autorità indipendenti hanno il compito di rilevare, monitorare e prevenire i fenomeni di discriminazione. Per esempio la Commissione Europea ha il compito di pubblicare annualmente una relazione sui progressi compiuti dagli Stati membri in merito alla direttive di uguale trattamento. In alcuni casi le autorità anti-discriminatorie hanno anche il potere di condurre investigazioni ed inchieste, fino ad avere funzioni quasi giuridiche e arrivando persino a inviare raccomandazioni ai governi.

1.1.2 La discriminazione diretta e indiretta

La politica anti-discriminatoria proibisce principalmente due tipi di comportamenti:

- Discriminazione diretta: quando certe regole o procedure trattano *esplicitamente* una persona in modo più sfavorevole di come un'altra è trattata o dovrebbe essere trattata, trovandosi in una stessa situazione. É chiamata anche discriminazione sistematica o trattamento non paritario.
- Discriminazione indiretta: quando certe regole o procedure o richieste apparentemente neutre risultano *implicitamente* avere un impatto ingiusto e discriminatorio su un gruppo protetto.

La discriminazione diretta è chiara e volontaria verso un gruppo sulla base di un loro visibile trattamento. Ad esempio risulta essere discriminazione diretta il rifiutare di affittare un appartamento ad una famiglia straniera, anche se è in grado di provare

la propria affidabilità economica; oppure il rifiutarsi di servire persone romane in un ristorante. Questi atti discriminatori diretti sono però difficili da dimostrare in quanto coloro che sono stati discriminati devono dimostrare la volontà stessa di essere state discriminate o ci dovrebbero essere dei testimoni forniti da coloro che sono stati discriminati.

Meno evidente è invece la discriminazione indiretta o strutturale: apparentemente si verifica tramite pratiche neutrali correlate ad attributi personali come razza, genere o altri gradi discriminatori, i cui effetti mettono però un gruppo in una posizione avversa e sfavorevole. Per esempio stabilire che uomini e donne devono avere un'altezza minima di 170 cm per entrare nelle forze armate è un parametro apparentemente neutro poiché si applica indistintamente a entrambi i sessi, ma pone le donne in una situazione di svantaggio di fatto perché l'altezza media delle donne è inferiore ai 170 cm, mentre per gli uomini è superiore. In realtà si tratta di un esempio di discriminazione indiretta purtroppo rilevato nell'ordinamento italiano e sostituito con la previsione di altezze minime diverse per uomini e donne. Si veda un altro esempio di discriminazione indiretta: la giurisprudenza degli USA stabilisce per lavori non specializzati e non richiedenti un grado di alfabetizzazione che i candidati abbiano il titolo di diploma; è chiaro che ciò produce effetti discriminatori nei confronti della popolazione di colore che si vuole potenzialmente candidare.

1.1.3 Le cause della discriminazione

Si possono classificare tre cause della discriminazione:

- Il pregiudizio
- Il razzismo razionale
- La discriminazione non intenzionale

Il termine *pregiudizio* significa letteralmente giudicare un'altra persona o un gruppo a priori⁵. Il termine pregiudizio si lega alla discriminazione nel senso che riguarda un ingiusto e non razionale atteggiamento negativo nei confronti di un gruppo protetto. Tale atteggiamento negativo è prodotto da stereotipi negativi,

⁵ Romei e Ruggieri, *A multidisciplinary survey on discrimination data analysis*, p. 3.

sensazioni negative o distanze sociali che si rafforzano a vicenda. Con il termine distanza sociale ci riferiamo ad scarsa familiarità con i membri di un gruppo protetto. Le sensazioni negative sono delle componenti affettive di date credenze, mentre con il termine stereotipi negativi ci riferiamo a credenze popolari proprie di tutti o della maggior parte dei membri del gruppo protetto. Il circolo vizioso della discriminazione nasce da una situazione nella quale il pregiudizio fa sì che il gruppo protetto sia socialmente svantaggiato; questo è interpretato come la prova che il gruppo sia inferiore e da qui si crea un ulteriore pregiudizio, aumentando la distanza sociale, rafforzando gli stereotipi negativi e legittimando le sensazioni negative.

Il *razzismo* è il risultato dell'atto di pensare razionalmente; una forma del razzismo razionale è la discriminazione statistica, ossia quando la mancanza della conoscenza sulle abilità di un individuo è compensata da una conoscenza a priori delle *performance* comuni del gruppo o della categoria alla quale l'individuo appartiene. Questo è considerato una sorta di pregiudizio razionale stereotipato. Un esempio può essere quando un impiegato presagisce un'impressione negativa sui suoi affari a causa del pregiudizio dei suoi clienti nei confronti di impiegati appartenenti a gruppi protetti.

Infine la *discriminazione non intenzionale* si ha non a causa di decisioni intenzionalmente maligne, ma a causa di una mancanza di consapevolezza sugli effetti decisionali. È una forma di discriminazione implicita e molte volte inconsapevole; questo è il caso di procedure o pratiche incorrette.

1.1.4 Le azioni affermative, la discriminazione inversa e i favoritismi

Le *azioni affermative*, chiamate anche azioni positive, sono quelle azioni fatte allo scopo di superare e compensare la discriminazione presente e passata, offrendo ai discriminati le opportunità che erano state negate loro. Queste azioni vanno dal mero incoraggiamento a quei gruppi, al trattamento preferenziale a favore di quei gruppi. Per esempio anche in ambito universitario a volte sono state fatte politiche di ammissione preferenziali nei confronti di donne o particolari categorie di candidati.

La *discriminazione inversa* si definisce come lo svantaggio che soffrono i non membri del gruppo protetto, come risultato dell'azione positiva, favorendo così una

discriminazione a favore di un altro gruppo. Per esempio nel caso di politiche di ammissioni preferenziali verso le donne è vero che ne traggono beneficio le donne, ma i restanti uomini sono chiaramente sfavoriti e subiscono così una discriminazione inversa.

Il *favoritismo*, chiamato anche nepotismo, avviene quando qualcuno sembra essere trattato in modo migliore rispetto ad un altro per ragioni non relazionate al merito individuale o alle azioni affermative. La differenza tra il nepotismo e le azioni affermative si nasconde nel gruppo che è favorito: nel caso di azioni affermative il gruppo è storicamente svantaggiato e la pratica è suggerita o richiesta dalla legge; nel favoritismo invece il gruppo è favorito per ragioni non supportate da regole o legislazioni, come nel caso della ragione consanguinea. Il favoritismo può anche essere attuato nei confronti di membri del gruppo protetto, come nel caso delle ammissioni alle università di persone di colore solo a causa dei loro risultati migliori nelle competitività sportive⁶.

1.2 La discriminazione in ambito legale

Da un punto di vista legale la discriminazione nasce dall'applicazione di regole o pratiche differenti nei confronti di situazioni simili, oppure dall'applicazione di regole o pratiche simili nei confronti di situazioni differenti.

Affinché si possa parlare di discriminazione, è quindi necessario che il comportamento non sia motivato da altri fattori, ma solo dall'appartenenza della persona o delle persone discriminate a detta categoria e che tale appartenenza non sia oggettivamente rilevante. Infatti, se discriminazione significa fondamentalmente “distinzione”, non tutte le distinzioni sono ingiustificate e quindi discriminatorie e come tali non ammesse dalla legge. Un regista che cerchi un'attrice per la parte di una donna, distinguerà tra attori uomini e attrici donne; il suo comportamento è però giustificato da fattori oggettivi e rilevanti rispetto alla scelta e quindi non è discriminatorio. Come si è già detto quelle categorie tradizionalmente soggette ad essere discriminate sono citate come gruppi protetti da leggi nazionali e internazionali sui diritti umani.

⁶ Romei e Ruggieri, *A multidisciplinary survey on discrimination data analysis*, p. 4.

1.2.1 Le prove statistiche dei casi discriminatori

La discriminazione sociale nelle pratiche economiche è un problema che, nel campo giuridico, vede la ricerca di regolamentazione da ormai oltre cinquant'anni. Le prove statistiche sono presenti in tutti i tipi di cause con regolamentazioni dei governi che citano la loro adozione. Per esempio, *the Reference Manual on Scientific Evidence*⁷, il cui scopo è quello di assistere i giudici federali degli USA a riconoscere le caratteristiche e la ragionevolezza della scienza come rilevante nella causa, include un capitolo di *Kaye e Freedman* sull'analisi statistica. Come osservato da *Gastwirth* l'oggettività della scienza e della legge possono talvolta divergere: un rigoroso metodo scientifico da un lato e un'avversa natura del sistema legale da un altro lato. La prova statistica dei casi discriminatori è considerata spesso come persuasiva, specialmente nel settore lavorativo. Il libro di *Finkelstein e Levin*⁸ (2001) cita numerosi casi di legge che illustrano metodi statistici di interpretazione della disparità: il passare/bocciare test di abilità da parte di afro-americani o ispano-americani; il salario tra impiegati maschi /femmine o bianchi/neri. Il principio legale della rappresentazione del gruppo si rifà prontamente all'analisi statistica della tabella sottostante (“v. tab. 1”). Il ricorso all'analisi statistica dei dati è riconosciuto come fondamentale sia nella legge comunitaria che nella legge nazionali di ogni paese membro. Un principio legale generale è quello di considerare il gruppo sotto rappresentato, ottenente un beneficio, come una misura quantitativa della discriminazione (indiretta) nei confronti di un gruppo protetto. La questione principale nei giudizi legali consiste nel determinare l'attinente popolazione, chiamata anche fonte di comparazione, il grado di discriminazione e la soglia che costituisce la prima prove evidente di discriminazione.

| | benefit | | |
|-------------|---------|---------|-------|
| | denied | granted | |
| protected | a | b | n_1 |
| unprotected | c | d | n_2 |
| | m_1 | m_2 | |

⁷ Romei e Ruggieri, *A multidisciplinary survey on discrimination data analysis*, p. 8.

⁸ Romei e Ruggieri, *A multidisciplinary survey on discrimination data analysis*, p. 8.

$p1 = a/n1 =$ benefici negati al gruppo protetto

$p2 = c/n2 =$ benefici negati al gruppo non protetto

$p = m1/n =$ il campione nell'insieme

Le differenze e le parti di queste proporzioni possono essere adottate come una controparte formale del principio legale del gruppo sotto rappresentato. Si chiamano in statistica:

Risk difference (RD) = $p1 - p2$

Relative risk / risk ratio (RR) = $p1 / p2$

Relative chance / selection rate (RC) = $(1-p1) / (1-p2)$

Odds ratio (OR) = $p1(1-p2) / (p2 (1-p1))$

La popolazione attinente: significa decidere dove scrivere n nella tabella 1. Ci sono molte possibilità e la scelta dipende dalla casistica legale nella quale ci si trova: il totale della popolazione di uno Stato, nel caso di un giudice che valuta un caso discriminatorio; una parte di persone candidate ad una posizione lavorativa, nel caso di una discriminazione lavorativa nel settore della forza lavoro. La scelta di un *range* di popolazione è decisiva nel caso di discriminazione indiretta, poiché influenza la grandezza del gruppo sotto rappresentato.

Per misurare il grado di discriminazione ci sono vari metodi che ogni autorità utilizza. La legislazione degli Usa preferisce il metodo del *risk difference*, la Corte di Giustizia dell'Unione Europea utilizza il metodo del *risk ratio*⁹.

1.2.2 La legislazione sulla discriminazione

Questa sezione vuole essere in parte una raccolta critica di leggi, organizzata per provenienza e campo di applicazione, per mostrare quanto il problema sia reale, essendo stato affrontato praticamente in ogni suo aspetto. Questo tema è presente anche nella Dichiarazione Universale dei Diritti dell'Uomo ed nella Carta dei Diritti Fondamentali dell'Unione Europea .

⁹ Romei e Ruggieri, *A multidisciplinary survey on discrimination data analysis*, p. 6.

➤ Fonte: ONU

Dichiarazione Universale dei Diritti dell'Uomo (Assemblea Generale delle Nazioni Unite, 10 Dicembre 1948)

Articolo 2:

“Ad ogni individuo spettano tutti i diritti e tutte le libertà enunciati nella presente Dichiarazione, senza distinzione alcuna, per ragioni di razza, di colore, di sesso, di lingua, di religione, di opinione politica o di altro genere, di origine nazionale o sociale, di ricchezza di nascita o di altra condizione”.

Articolo 7:

“Tutti sono eguali dinanzi alla legge e hanno diritto, senza alcuna discriminazione, ad un'eguale tutela da parte della legge. Tutti hanno diritto ad un'eguale tutela contro ogni discriminazione che violi la presente Dichiarazione come contro qualsiasi incitamento a tale discriminazione”.

I due articoli hanno valenza generale, come ci si aspetta dal Documento, e devono dare semplicemente le linee guida ai legislatori dei Paesi membri.

➤ Fonte: UE

Carta dei Diritti Fondamentali dell'Unione Europea, Capo III (Parlamento, Consiglio e Commissione Europea, Nizza 7 Dicembre 2000).

Articolo 21: Non discriminazione

1. *“E vietata qualsiasi forma di discriminazione fondata, in particolare, sul sesso, la razza, il colore della pelle o l'origine etnica o sociale, le caratteristiche genetiche, la lingua, la religione o le convinzioni personali, le opinioni politiche o di qualsiasi altra natura, l'appartenenza ad una minoranza nazionale, il patrimonio, la nascita, gli handicap, l'età o le tendenze sessuali”.*

2. *“Nell’ambito d’applicazione del trattato che istituisce la Comunità Europea e del trattato sull’Unione Europea è vietata qualsiasi discriminazione fondata sulla cittadinanza, fatte salve le disposizioni particolari contenute nei trattati stessi”.*

Articolo 22: Diversità Culturale, religiosa e linguistica

“L’Unione rispetta la diversità culturale, religiosa e linguistica”.

Articolo 23: Parità tra uomini e donne

“La parità tra uomini e donne deve essere assicurata in tutti i campi, compreso le materie di occupazione, di lavoro e di retribuzione. Il principio della parità non osta al mantenimento o all’adozione di misure che prevedano vantaggi specifici a favore del sesso sotto rappresentato”.

Articolo 26: Inserimento dei disabili

“L’Unione riconosce e rispetta il diritto dei disabili di beneficiare di misure intese a garantirne l’autonomia, l’inserimento sociale e professionale e la partecipazione alla vita della comunità”.

La Carta riprende rigorosamente i principi dell’Articolo 22 della Dichiarazione con l’aggiunta di alcuni caratteri molto interessanti, quali l’età, le tendenze sessuali e gli handicap in termine di discriminazione. Evidentemente questi caratteri si sono manifestati come oggetto di discriminazione in un momento successivo al 1948, come è facile immaginare.

➤ Fonte: diritto civile italiano

Legge 40/98, Articolo 41, Comma 1: Discriminazione per motivi razziali, etnici, nazionali o religiosi

Capo 1: *“Ai fini del presente capo, costituisce discriminazione ogni comportamento che, direttamente o indirettamente, comporti una distinzione, esclusione, restrizione o preferenza basata sulla razza, il colore, l’ascendenza o l’origine nazionale o*

etnica, le convinzioni e le pratiche religiose, e che abbia lo scopo o l'effetto di distruggere o di compromettere il riconoscimento, il godimento o l'esercizio, in condizioni di parità, dei diritti umani e delle libertà fondamentali in campo politico economico, sociale e culturale e in ogni altro settore della vita pubblica".

Capo 2: *"In ogni caso compie atto di discriminazione:*

1. il pubblico ufficiale o la persona incaricata di pubblico servizio o la persona esercente un servizio di pubblica necessità che nell'esercizio delle sue funzioni compia od ometta atti nei riguardi di un cittadino straniero che, soltanto a causa della sua condizione di straniero o di appartenente ad una determinata razza, religione, etnia o nazionalità, lo discriminino ingiustamente;

2. chiunque imponga condizioni più svantaggiose o si rifiuti di fornire beni o servizi offerti al pubblico ad uno straniero soltanto a causa della sua condizione di straniero o di appartenente ad una determinata razza, religione, etnia o nazionalità;

3. chiunque illegittimamente imponga condizioni più svantaggiose o si rifiuti di fornire l'accesso all'occupazione, all'alloggio, all'istruzione alla formazione e ai servizi sociali e socio-assistenziali allo straniero regolarmente soggiornante in Italia soltanto in ragione della sua condizione di straniero o di appartenente ad una determinata razza, religione, etnia o nazionalità;

4. chiunque impedisca, mediante azioni od omissioni, l'esercizio di un'attività economica legittimamente intrapresa da uno straniero regolarmente soggiornante in Italia, soltanto in ragione della sua condizione di straniero o di appartenente ad una determinata razza, confessione religiosa, etnia o nazionalità;

5. il datore di lavoro o i suoi preposti i quali, ai sensi dell'articolo 15 della legge 20 maggio 1970, n. 300, come modificata e integrata dalla legge 9 dicembre 1977, n. 903, e dalla legge 11 maggio 1990, n. 108, compiano qualsiasi atto o comportamento che produca un effetto pregiudizievole discriminando, anche indirettamente, i lavoratori in ragione della loro appartenenza ad una razza, ad un

gruppo etnico o linguistico, ad una confessione religiosa, ad una cittadinanza. Costituisce discriminazione indiretta ogni trattamento pregiudizievole conseguente all'adozione di criteri che svantaggino in modo proporzionalmente maggiore i lavoratori appartenenti ad una determinata razza, ad un determinato gruppo etnico o linguistico, ad una determinata confessione religiosa o ad una cittadinanza e riguardino requisiti non essenziali allo svolgimento dell'attività lavorativa".

Capo 3: *"Il presente articolo e l'articolo 42 si applicano anche agli atti xenofobi, razzisti o discriminatori compiuti nei confronti dei cittadini italiani, di apolidi e di cittadini di altri Stati membri dell'Unione europea presenti in Italia".*

Con questa legge del 1998 si allarga il campo di applicazione della precedente legge Mancino, riferendosi anche alle pratiche economiche, anche se l'oggetto di discriminazione resta sempre quello etnico-religioso.

➤ D. LGS. 216/03, Articolo 2

Comma 1: *"Ai fini del presente decreto e salvo quanto disposto dall'articolo 3, commi da 3 a 6, per principio di parità di trattamento si intende l'assenza di qualsiasi discriminazione diretta o indiretta a causa della religione, delle convinzioni personali, degli handicap, dell'età o dell'orientamento sessuale. Tale principio comporta che non sia praticata alcuna discriminazione diretta o indiretta, cos'è come di seguito definite:*

a) discriminazione diretta quando, per religione, per convinzioni personali, per handicap, per età o per orientamento sessuale, una persona è trattata meno e favorevolmente di quanto sia, sia stata o sarebbe trattata un'altra in una situazione analoga;

b) discriminazione indiretta quando una disposizione, un criterio, una prassi, un atto, un patto o un comportamento apparentemente neutri possono mettere le persone che professano una determinata religione o ideologia di altra natura, le persone portatrici di handicap, le persone di una particolare età o di un orientamento

sessuale in una situazione di particolare svantaggio rispetto ad altre persone”.

Comma 2: *“E fatto salvo il disposto dell’articolo 43, commi 1 e 2 del testo unico delle disposizioni concernenti la disciplina dell’immigrazione e norme sulla condizione dello straniero, approvato con decreto legislativo 25 luglio 1998, n. 286”.*

Comma 3: *“Sono, altresì considerate come discriminazioni, ai sensi del comma 1, anche le molestie ovvero quei comportamenti indesiderati, posti in essere per uno dei motivi di cui all’articolo 1, aventi lo scopo o l’effetto di violare la dignità di una persona e di creare un clima intimidatorio, ostile, degradante, umiliante od offensivo”.*

Comma 4: *“L’ordine di discriminare persone a causa della religione, delle convinzioni personali, dell’handicap, dell’età o dell’orientamento sessuale è considerata una discriminazione ai sensi del comma 1”.*

Questo decreto legislativo fornisce la definizione di discriminazione indiretta già citata e spiegata (“v. sottopar. 1.1.2). Inoltre, allarga il campo di applicazione anche alla discriminazione sulla base di handicap, età ed orientamento sessuale, mai considerate prima d’ora.

1.3 La discriminazione in ambito economico

In questo paragrafo si tratterà della discriminazione in ambito economico, seppur risulta evidente che tale scelta implichi una riduzione del campo di studio: basti pensare ad una situazione in cui un individuo riceve un trattamento economico giusto ma viene deriso, schernito o turbato psicologicamente, in questo caso l’analisi economica non coglie alcun tipo di discriminazione. In particolare ci si soffermerà sulla discriminazione di genere.

Oggi giorno nel mercato del lavoro, trattamenti differenti tra gruppi di lavoratori possono essere misurati attraverso la loro remunerazione, il rispettivo grado di partecipazione nella forza lavoro o il rispettivo grado di segregazione di lavoratori in

una specifica occupazione o industria. Molti sondaggi e indagini pubbliche raccolgono quotidianamente dati sulle caratteristiche demografiche e attitudini dei residenti (ad esempio negli U.S., il *General Social Survey*), sulla distribuzione della forza lavoro nel mercato del lavoro (ad esempio il *Current Population Survey*), e così via. Differenze di genere e razziali sono state sempre rilevate: negli U.S.A. la percentuale di disoccupati di età compresa fra i 25 e 54 anni nel 2010 degli uomini di colore era del 16.3 %, rispetto a 8,5% dei bianchi.

1.3.1 I modelli economici

All'incirca dalla metà degli anni sessanta, in corrispondenza con l'entrata in vigore negli Stati Uniti d'America di una prima serie di leggi federali contenenti divieti espliciti di discriminazione nei rapporti di lavoro (*Equal Pay Act* 1963, titolo VII del *Civil Rights Act* 1964, *Age Discrimination in Employment Act* 1967), gli economisti statunitensi hanno incominciato a considerare la questione delle pratiche discriminatorie poste in essere dalle imprese e degli effetti delle norme antidiscriminatorie come una delle materie più rilevanti nello studio del funzionamento del mercato del lavoro. Per spiegare il comportamento discriminatorio dei datori di lavoro sono stati elaborati numerosi modelli, da ciascuno dei quali possono trarsi indicazioni teoriche parzialmente differenti circa gli effetti della legislazione vigente e circa le tecniche legislative più efficaci per contrastare il fenomeno. Infatti tecniche di ricerca empirica hanno utilizzato inferenze statistiche allo scopo sia di raccogliere i dati, sia di verificare le conseguenze predette da un modello economico teorico e anche di accertare il contributo di differenti tipi di discriminazione verso trattamenti differenziati nel mercato del lavoro. Le principali tecniche di analisi di dati adottate includono verifiche statistiche su modelli di regressione lineare generalizzati e modelli econometrici.

Nella letteratura economica sono stati considerati due maggiori modelli teorici:

- *Taste-based-discrimination*: non ha fondamenta razionali o economiche, ma si basa solo su un sensazione personale nata da un pregiudizio nei confronti di un gruppo protetto.
- *Statistical-discrimination*: ha una base scientifica ed economica e parte dall'assunzione che gli impiegati non possono stimare perfettamente la

produttività di ogni lavoratore al momento dell'assunzione.

Le imperfezioni del mercato ci danno un incentivo a considerare più facilmente le caratteristiche osservabili, come il genere / razza per stimare la produttività sperata, calcolata in base alla conoscenza a priori della produttività media del gruppo al quale loro stessi appartengono. Le remunerazioni sono alla base della stimata produttività media del gruppo e non alla base della produttività di ogni singola persona.

Il modello di *taste-based-discrimination* è stato ideato da Backer¹⁰ e distingue tra la discriminazione dell'imprenditori, la discriminazione del cliente e la discriminazione del collaboratore. Il modello tende a dimostrare che gli imprenditori caratterizzati da un particolare "gusto per la discriminazione" (*taste for discrimination*) offrono retribuzioni inferiori ai lavoratori appartenenti al gruppo discriminato; così sono destinati a conseguire, in condizioni di equilibrio di mercato, profitti inferiori rispetto a quelli conseguiti dagli altri imprenditori, a parità di ogni altra condizione. Quindi l'imprenditore che discrimina in base alla razza o al sesso del lavoratore accetta una riduzione dei profitti pur di soddisfare le proprie preferenze razziali o di genere.

Nel modello di *statistical-discrimination* invece tale comportamento è funzionale a una massimizzazione dei profitti. Infatti l'imprenditore utilizza l'appartenenza del lavoratore a un determinato gruppo come indice di una caratteristica probabile del lavoratore stesso, rilevante ai fini della previsione della sua produttività futura o del suo costo futuro per l'azienda. Ad esempio, negli Stati Uniti è statisticamente vero che i bianchi frequentano mediamente scuole di qualità migliore rispetto ai neri a parità di titolo di studio e di altre condizioni; quindi il fatto che il lavoratore sia bianco significa una sua probabile produttività futura maggiore. Allo stesso modo, è statisticamente vero, e non soltanto negli Stati Uniti, che le donne hanno mediamente tassi di assenza dal lavoro più elevati degli uomini, con corrispondente riduzione della loro produttività media. Oppure ancora: le donne hanno mediamente una aspettativa di vita maggiore degli uomini, con un conseguente maggior costo per l'azienda che eroghi agli ex- dipendenti un trattamento pensionistico integrativo.

Questi e altri modelli consentono di spiegare il persistere dei comportamenti discriminatori, pur in un mercato perfettamente concorrenziale, e sottolineano la necessità di un intervento dello Stato volto a correggere il fenomeno. In particolare

¹⁰ Romei e Ruggieri, *Discrimination Data Analysis: A Multi-disciplinary Bibliography*, p. 5.

però alcuni effetti controproducenti di norme che vietano la differenziazione di trattamento, intervengono imponendo parità retributiva in situazioni nelle quali è ancora in atto il circolo vizioso fra discriminazione e minore investimento dei discriminati sul proprio capitale umano, ritorcendo così il divieto di differenziazione contro coloro che dovrebbero esserne protetti. Si è inoltre osservato che le norme antidiscriminatorie hanno sempre dimostrato di poter incidere sul comportamento degli imprenditori molto di più in fase di svolgimento o risoluzione del contratto di lavoro, che in fase di assunzione; il risultato è che al momento della scelta del lavoratore da assumere, il temuto rischio di una futura controversia finisce col costituire un disincentivo all'assunzione del lavoratore appartenente al gruppo svantaggiato.

1.4 I pregiudizi di genere nella ricerca scientifica

Nei primi anni novanta sono state fatte delle scoperte, nei settori degli studi sociali economici, riguardanti la discriminazione nei confronti delle donne, grazie a ricerche che sono state portate avanti da *S. J. Ceci* e *W. M. Williams*¹¹. Tuttavia alcuni studiosi hanno pubblicato altri articoli recenti principalmente incentrati sull'argomento della discriminazione nei confronti del sesso debole per selezioni o finanziamenti. Un primo filone è condotto da *Bornmann*¹² ed alcuni suoi colleghi, i quali hanno analizzato più di mille domande per borse di studio pre e post dottorato in Germania, con lo scopo di rilevare favoritismi di genere nelle procedure di selezione.

In un altro studio i soliti autori si sono concentrati su un'analisi di ricerca letteraria sulle sovvenzioni in base al sesso. La base di dati contiene riferimenti, pubblicati tra il 1987 e 2005, su riviste di ricerca, integrate con una ricerca delle pubblicazioni per ogni autore. Vennero analizzate un totale di sessantasei procedure di *peer-review*, raccolte su un totale di 353,725 richieste. Da un punto di vista metodologico i suddetti studi utilizzano, come strumento base l'analisi di regressione e in particolare utilizza regressioni ordinali multiple per stimare il numero di referenze ricevute da ogni ricercatore. Le variabili indipendenti includono l'età del candidato, il titolo di studio, il numero di pubblicazioni, la nazionalità, il genere.

11 Romei e altri, *Discrimination Discovery in Scientific Project Evaluation: A Case Study*, p. 3.

12 Romei e altri, *Discrimination Discovery in Scientific Project Evaluation: A Case Study*, p. 3.

*H. W Marsh*¹³ ha continuato il filone di studi condotto da *Bornmann*, usando però diversi approcci di meta-analisi di regressione; infatti mentre *Bornmann* utilizza un approccio a tre livelli (studi, *peer review*, singole richieste), Marsh sfrutta gli effetti fissi e casuali sulla meta-analisi di regressione. I risultati ottenuti sono però abbastanza contrastanti; *Bornmann* ha rilevato piccole differenze di genere a favore degli uomini, mentre *Marsh* ha rifiutato questa tesi, non trovando significative differenze di genere.

Altri studi sono stati condotti da *Brouns*, che ha investigato su come ottenere borse di studio prestigiose in Germania. I risultati dimostrano che nelle discipline fisiche e matematiche, nelle quali le donne risaltano, la qualità delle produttività scientifica è un fattore principale per ottenere le borse di studio. Nelle discipline biologiche e geologiche, nelle quali invece il sesso maschile prevale, c'è una forte correlazione tra la decisione di concedere una borsa di studio e la rapidità del dottorato di ricerca.

13 Romei ed altri, *Discrimination Discovery in Scientific Project Evaluation: A Case Study*, p. 4.

2. La base di dati

2.1 Il *database* del FIRB

Nel precedente capitolo si è spiegato e passato in rassegna il concetto di discriminazione, in particolare in ambito legale ed economico. Inoltre si sono illustrate alcune ricerche scientifiche sui pregiudizi nei confronti delle donne in ambiti di ricerca. In questa sezione si affronteranno le particolarità del *database* FIRB, spiegando nel particolare la sua struttura e il suo tipo di gestione. Inoltre si illustreranno le modalità di finanziamento per i progetti.

La base di dati FIRB è disponibile tramite un *Oracle relationship database*, formato da ben quindici tabelle. Questo *database* tiene traccia di tutti i progetti di ricerca inviati nell'anno 2008¹ e di tutti i dati relativi alle attività di ricerca. In questo caso tramite un processo di *data warehousing* che comprende metodi, tecnologie e strumenti d'aiuto, si è giunti ad una sintesi dei dati più rilevanti per questo tipo di analisi. Nei processi di *data warehousing* si raggiunge la sintesi dei dati al livello del *warehouse*, ovvero quando le informazioni vengono raccolte in un singolo "contenitore" centralizzato logicamente: il *data warehouse*. Esso è usato, in questa casistica, come sorgente per costruire un *data mart*, che ne costituisce una parziale replica dei dati FIRB. Quest'ultimo contiene l'insieme delle informazioni più rilevanti per specifiche analisi, verso il settore della *discrimination discovery* fra uomo e donna. Quindi i *data mart*, che si andranno a verificare, possono essere visti come delle basi di dati locali che replicano in sintesi il *database* primario per effettuare verifiche nell'area della *discrimination discovery*.

2.2 Le basi di dati

Le basi di dati ispezionate sono due e si distinguono per due linee di ricerca, giovani e strutturati.

- Giovani: è il personale con un contratto a tempo determinato che lavora nel dipartimento / istituto. Hanno un anno di nascita compreso tra il 1975 e 1982.
- Strutturati: sono i ricercatori e i professori con un contratto a tempo

¹ Romei ed altri, *Discrimination Discovery in Scientific Project Evaluation: A Case Study*, p. 7.

indeterminato e hanno un anno di nascita compreso tra il 1969 e 1982.

Conseguentemente definiremo, per motivi pratici, con la sigla L1 il *data mart* dei giovani, mentre con la sigla L2 il *data mart* degli strutturati. Entrambi le basi di dati hanno la solita struttura e differiscono solo per numero di enuple. Questi *database* sono stati gestiti ed implementati tramite uno *spreadsheet*: Calc.

2.2.1 La descrizione delle basi di dati

I *data mart* tengono traccia delle seguenti informazioni:

- a) *Informazioni personali e professionali sui coordinatori di progetto*
- b) *Costi relativi del progetto*
- c) *Proprietà del progetto*
- d) *Valutazione del progetto*

Si possono classificare i seguenti attributi in quattro classi, che sono specificate nella tabella 2.1.

- a) Le informazioni personali e professionali relative al coordinatore includono: il sesso, il numero di pubblicazioni, la media autori, la loro qualifica, la zona e città di provenienza, l'organizzazione per la quale lavorano ed infine il numero di donne con le quali lavorano al progetto.
- b) I costi del progetto sono composti da: il costo totale di ogni progetto, i fondi richiesti, la percentuale di richiesta rispetto al costo totale del progetto, il numero di giovani ricercatori associati a quel progetto, il costo dei giovani ricercatori, la percentuale del costo dei giovani ricercatori in base al costo del progetto totale, il numero di ricercatori di fama internazionale, il costo dei ricercatori di fama internazionale e la percentuale del costo dei ricercatori di chiara fama internazionale rispetto al costo del progetto totale.
- c) Un totale di nove attributi descrive i domini dei settori di ricerca, secondo la

classificazione europea del Consiglio di Ricerca². Questa classificazione è organizzata secondo tre livelli di gerarchia. Il livello più alto comprende tre settori principali: scienze sociali Lettere e Filosofia (SH), scienze fisiche e ingegneristiche (PE), e scienze della vita (LC). Il secondo e terzo livello comprendono rispettivamente un totale di 25 e 3,792 materie specifiche.

d) La valutazione del progetto seguendo i rispettivi valori di criterio {1, 2, 3, 4}.

- *Scientific relevance (score 0-8)*
- *Impact of the project (score 0-7)*
- *Scientific and technical value (score 0-15)*
- *Quality of partnership (score 0-10)*

² <http://erc.europa.eu/>

| Nome | Tipo | Descrizione |
|-----------------------------------------------------------------------------|-------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>CODE_A</i> | <i>Integer</i> | <i>Identificatore progetto</i> |
| Informazioni personali e professionali sul coordinatore del progetto | | |
| <i>SESSO</i> | <i>Enumerated</i> | <i>Sesso del coordinatore</i> |
| <i>ANNO_NASCITA'</i> | <i>Integer</i> | <i>Anno nascita del coordinatore</i> |
| <i>CITTA</i> | <i>String</i> | <i>Città del coordinatore</i> |
| <i>ZONA</i> | <i>String</i> | <i>Zona del coordinatore</i> |
| <i>TIPO_ENTE</i> | <i>String</i> | <i>Istituto appartenenza del coordinatore</i> |
| <i>QUALIFICA</i> | <i>Enumerated</i> | <i>Qualifica del coordinatore: "Ric astronomo non confermato", "Dir EPR", "Primo ric", "Prof Straordinario", "Ric", "Ric confermato", "Ric non confermato", "Dottore di Ricerca", "Ric astronomo confermato", "Tecnologo", "Prof Associato confermato", "Prof Associato non confermato"</i> |
| <i>NUM_PUBB</i> | <i>Integer</i> | <i>Numero pubblicazioni coordinatore</i> |
| <i>AVG_NUM_AUTORI</i> | <i>Integer</i> | <i>Media pubblicazioni coordinatore</i> |
| <i>NUM_DONNE</i> | <i>Integer</i> | <i>Numero di donne nel progetto</i> |
| Costi progetto | | |
| <i>CONTRIBUTO_MIUR</i> | <i>Integer</i> | <i>Contributo Miur (>0 per prog. Finanziati; =0 altrimenti)</i> |

| | | |
|----------------------------------|----------------|-----------------------------------------------------------------------------------------------------|
| <i>CONTRIBUTO_MIUR_PERC</i> | <i>Integer</i> | <i>Indica la quantità in percentuale del contributo Miur rispetto al costo totale</i> |
| <i>COSTO_TOT</i> | <i>Real</i> | <i>Costo totale del progetto</i> |
| <i>FIN_RIC</i> | <i>Real</i> | <i>Finanziamento richiesto</i> |
| <i>FIN_RIC_PERC</i> | <i>Integer</i> | <i>Finanziamento richiesto in percentuale rispetto al costo totale del progetto</i> |
| <i>GR_NUM</i> | <i>Integer</i> | <i>Numero giovani ricercatori</i> |
| <i>GR_COSTO</i> | <i>Real</i> | <i>Costi giovani ricercatori</i> |
| <i>GR_COSTO_PERC</i> | <i>Integer</i> | <i>Indica il costo dei giovani ricercatori in percentuale rispetto al costo totale del progetto</i> |
| <i>RCF_NUM</i> | <i>Integer</i> | <i>Numero di ricercatori di fama internazionale</i> |
| <i>RCF_COSTO</i> | <i>Real</i> | <i>Costi ricercatori di fama internazionale</i> |
| <i>RCF_COSTO_PERC</i> | <i>Integer</i> | <i>Indica il costo dei ricercatori famosi in percentuale rispetto al costo totale del progetto</i> |
| <i>Proprietà progetto</i> | | |
| <i>LINEA_INTERVENTO</i> | <i>Integer</i> | <i>Identificatore database L1 o L2</i> |
| <i>SETT_ID_1_LV_1</i> | <i>String</i> | <i>Identificatore del primo settore di riferimento</i> |
| <i>SETT_ID_1_LV_2</i> | <i>String</i> | <i>Identificatore del primo settore di riferimento</i> |
| <i>SETT_ID_1_LV_3</i> | <i>String</i> | <i>Identificatore del primo settore di riferimento</i> |

| | | |
|------------------------------------|-------------------|----------------------------------------------------------|
| <i>SETT_ID_2_LV_1</i> | <i>String</i> | <i>Identificatore del secondo settore di riferimento</i> |
| <i>SETT_ID_2_LV_2</i> | <i>String</i> | <i>Identificatore del primo settore di riferimento</i> |
| <i>SETT_ID_2_LV_3</i> | <i>String</i> | <i>Identificatore del primo settore di riferimento</i> |
| <i>SETT_ID_3_LV_1</i> | <i>String</i> | <i>Identificatore del terzo settore di riferimento</i> |
| <i>SETT_ID_3_LV_2</i> | <i>String</i> | <i>Identificatore del primo settore di riferimento</i> |
| <i>SETT_ID_3_LV_3</i> | <i>String</i> | <i>Identificatore del primo settore di riferimento</i> |
| <i>Valutazione progetto</i> | | |
| <i>FINANZIATO</i> | <i>Enumerated</i> | <i>Indica se un progetto è stato finanziato</i> |
| <i>AMMESSO</i> | <i>Enumerated</i> | <i>Indica se un progetto è stato ammesso</i> |
| <i>VAL_CRITERIO_1</i> | <i>Integer</i> | <i>Scientific relevance (score: 0 - 8)</i> |
| <i>VAL_CRITERIO_2</i> | <i>Integer</i> | <i>Impact (score 0 - 7)</i> |
| <i>VAL_CRITERIO_3</i> | <i>Integer</i> | <i>Scientific and technical value (score 0 - 15)</i> |
| <i>VAL_CRITERIO_4</i> | <i>Integer</i> | <i>Quality of partnership (score 0 - 10)</i> |

Tabella 1. Descrizione *data mart* del FIRB, L1 e L2.

2.3 Le modalità di finanziamento del progetto

L'accettazione finale al finanziamento di un progetto è sottoposto a due principali *step*: il primo è *international peer-review* e il secondo è l'*audit*. Se entrambi le due fasi hanno riscontro positivo il MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca) classifica l'insieme dei coordinatori e decide sia se il progetto sarà avviato, sia l'entità dei fondi concessi.

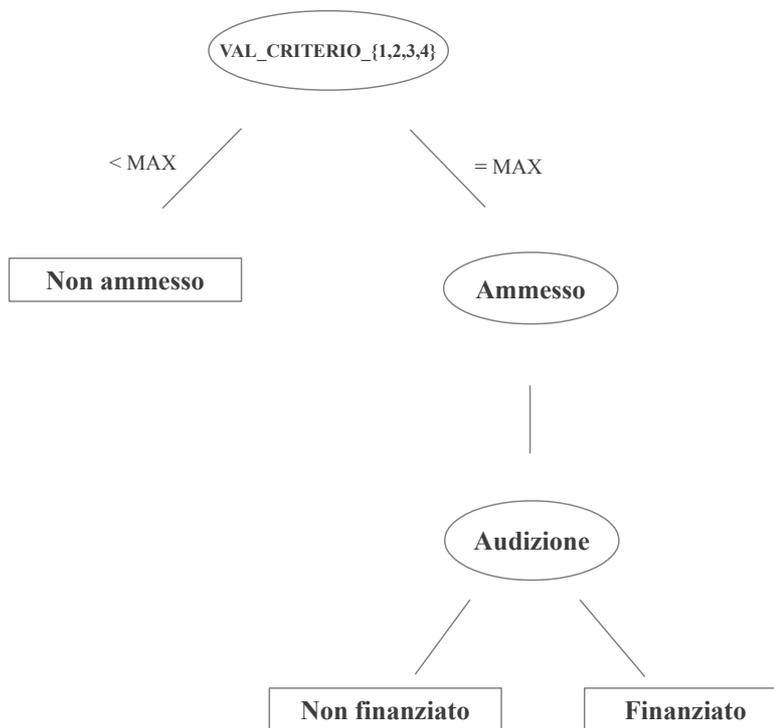
Nella prima fase ogni coordinatore è valutato in base a quattro criteri:

- *Scientific relevance (score 0-8)*
- *Impact of the project (score 0-7)*
- *Scientific and technical value (score 0-15)*
- *Quality of the partnership (score 0-10)*

Il progetto sarà ammesso solo nel caso in cui ottenga il massimo punteggio su tutte e quattro le valutazioni e conseguentemente avrà il diritto di passare alla fase successiva. In caso contrario non sarà ammesso e quindi respinto.

Nella seconda fase vengono esaminati i progetti ammessi tramite un audizione. Se l'audizione avrà un riscontro positivo il progetto sarà finanziato, altrimenti no. Questo modello è stato sintetizzato in base alle sue proprietà tramite il seguente processo decisionale (“v. fig. 2”).

Step 1:
International
peer review



Step 2:
Audition

Figura 2. Processo di finanziamento.

3. *Discrimination discovery*: approccio con l'analisi dei dati

3.1 Analisi dei dati per analisi discriminatorie

Lo studio della discriminazione attraverso l'utilizzo delle statistiche e delle scienze economiche è una linea di ricerca nata intorno ai primi anni cinquanta grazie agli studi di *Gary S. Becker*. Da allora, ovviamente, le teorie economiche e le ricerche empiriche si sono susseguite in modo forsennato da un lato facendo interessantissime scoperte, ma dall'altro anche trovando dei notevoli problemi alle loro applicazioni. Un grosso limite strumentale a questi studi, soprattutto nei primi anni, è stata la mancanza di mezzi tecnologici necessari per poter trattare la mole di dati che il contesto richiedeva. Oggigiorno invece, grazie alle tecnologie informatiche che si hanno a disposizione e alle applicazioni dell'analisi dei dati, le teorie in questo campo possono trovare uno sbocco pratico.

Questo capitolo illustra il lavoro svolto sui dati FIRB; in particolare dopo aver definito le proprietà dei due *data mart* (rispettivamente per L1 e L2) tramite delle analisi preliminari, illustrando le particolarità e le differenze, questo lavoro si concentrerà sull'analisi del processo di selezione dei progetti da finanziare secondo due fasi: analisi delle distribuzioni e analisi di regressione.

3.2 Analisi preliminari

In questa sezione si riportano una serie di *Pivot Table* (PT) che descrivono le basi statistiche sulle quali si è fondato il progetto di ricerca. Inoltre da ora in poi si considererà il sesso dei coordinatori come primaria risorsa di discriminazione. Si descriveranno le proprietà dei due *datamart* tramite l'analisi delle sue variabili più caratterizzanti e l'incrocio di quest'ultime grazie al cubo *OLAP* (*On Line Analytical Processing*).

I dati utilizzati per la costruzione del cubo si dividono in due differenti gruppi:
- misure: sono costituite dai dati contenuti nei fatti del *data mart*, si troveranno all'interno del cubo e costituiscono le informazioni che verranno lette.

- dimensioni: sono costituite dai dati contenuti nelle dimensioni dei fatti dei *data mart* e formano le dimensioni del cubo.

E' possibile estrarre informazioni dal cubo utilizzando due operazioni: la *SLICE* (taglia una fetta del cubo con una restrizione su una delle dimensioni) e la *DICE* (taglia un cubetto del cubo con una restrizione su una o più parti). Inoltre è possibile aggiungere informazioni al cubo grazie all'operazione di *ROLL-UP* che aumenta lo stato aggregativo del cubo. Il cubo *OLAP* ha una struttura che si adatta naturalmente al modello dimensionale del *datawarehouse* e ne amplia notevolmente le potenzialità, rendendolo navigabile in modo semplice e performante. Questa tecnica si applica tramite le PT che permettono di fare *OLAP* in calc.4.

Le prime analisi sono effettuate sulla variabile categorica *SESSO* che identifica il sesso dei coordinatori. Si crea una PT con dimensioni *SESSO* e dati *SESSO* in percentuale.

| | |
|-------------------------|-------------|
| SESSO | |
| F | 51% |
| M | 49% |
| Totale Risultato | 100% |

Tabella 1. L1

| | |
|-------------------------|-------------|
| SESSO | |
| F | 40% |
| M | 60% |
| Totale Risultato | 100% |

Tabella 2. L2

In L1 ci sono 1805 coordinatori dove il 49 % dei progetti è diretto da coordinatori uomini, mentre il 51 % da coordinatrici donne (“v.tab.1”). Rispettivamente in L2 abbiamo 1987 coordinatori in cui il 60 % è diretto da coordinatori uomini e il restante 39 % da coordinatrici donne (“v.tab.2”). Il risultato della tabella di L1 mostra che c'è un numero leggermente maggiore di progetti diretti da donne, rispetto agli uomini. Mentre in L2 notiamo una notevole differenza tra i progetti coordinati da donne e uomini.

Successivamente si eseguono due operazioni su quest'ultimi risultati:

1. Operazione di *slicing*, riducendo la dimensionalità per sesso;
2. Operazione di *roll-up*, aumentando l'aggregazione della tabella per la variabile nominale TIPO_ENTE, che descrive l'ente di appartenenza del coordinatore. Quest'ultima assume tre probabili valori:

- Altro
- CNR
- Università

| ContaNumeri - SESSO | SESSO | |
|-------------------------|-------------|-------------------------|
| TIPO_ENTE | F | Totale Risultato |
| Altro | 5% | 5% |
| CNR | 8% | 8% |
| Univ | 87% | 87% |
| Totale Risultato | 100% | 100% |

Tabella 3. L1 donne con operazioni *slicing* e *roll-up*

| ContaNumeri - SESSO | SESSO | |
|-------------------------|-------------|-------------------------|
| TIPO_ENTE | M | Totale Risultato |
| Altro | 9% | 9% |
| CNR | 9% | 9% |
| Univ | 83% | 83% |
| Totale Risultato | 100% | 100% |

Tabella 4. L1 uomini con operazioni *slicing* e *roll-up*

I risultati ottenuti per L1 mostrano, nel caso dei capi progetto donne, che il 5 % fa riferimento ad Altro, l'8 % al CNR ed infine l'87 % all'Università (“v.tab.3”). Mentre nel caso dei capi progetto uomini si evidenzia che il 9% appartiene ad Altro, il 9% al CNR e l'83 % all'Università (“v.tab.4”). Comparando i risultati delle tab.3 e tab.4, ottenute da L1, si nota che oltre l'80 % dei coordinatori per entrambi i sessi fa capo all'università, mentre il restante si divide fra Altro e CNR. Bisogna inoltre precisare che il numero di donne che fa capo all'Università è maggiore rispetto ai colleghi uomini.

Si guardino ora i risultati ottenuti per L2 :

| ContaNumeri - SESSO | SESSO | |
|-------------------------|-------------|-------------------------|
| TIPO_ENTE | F | Totale Risultato |
| Altro | 3% | 3% |
| CNR | 10% | 10% |
| Univ | 87% | 87% |
| Totale Risultato | 100% | 100% |

Tabella 5. L2 donne con operazioni *slicing* e *roll-up*

| ContaNumeri - SESSO | SESSO | |
|-------------------------|-------------|-------------------------|
| TIPO_ENTE | M | Totale Risultato |
| Altro | 5% | 5% |
| CNR | 7% | 7% |
| Univ | 88% | 88% |
| Totale Risultato | 100% | 100% |

Tabella 6. L2 uomini con operazioni *slicing* e *roll-up*

Dalla tab. 5 e tab. 6 si evince che oltre l'85 % dei coordinatori di entrambi i sessi fa riferimento all'Università e il restante al CNR o Altro.

In sintesi, valutando i risultati ottenuti per L1 e L2, si nota che oltre l'80 % dei coordinatori fa capo all'Università.

3.2.1 Analisi dei costi progetti

In questo passaggio si analizzano i costi dei progetti e le loro variabili più rilevanti. Di particolare interesse si mostra studiare il solito comportamento lungo il genere e lungo la linea di ricerca, rispetto ad una tendenza principale dei costi nel set di dati FIRB. Si prenderanno in esame:

- Variabili giovani ricercatori: GR_COSTO e GR_NUM
- Variabili ricercatori di fama internazionale: RCF_COSTO e RCF_NUM
- COSTO_TOTALE: costo totale dei progetti

Nel caso **GR_COSTO** si crea una PT con dimensione SESSO e con dati la media

di GR_COSTO e la media del GR_NUM.

| | | SESSO | | |
|------------------|--|---------|---------|------------------|
| Dati | | F | M | Totale Risultato |
| Media - GR_COSTO | | 239.809 | 262.438 | 250.860 |
| Media - GR_NUM | | 2 | 2 | 2 |

Tabella 7. L1 GR_COSTO

| | | SESSO | | |
|------------------|--|---------|---------|------------------|
| Dati | | F | M | Totale Risultato |
| Media - GR_COSTO | | 213.439 | 237.944 | 228.172 |
| Media - GR_NUM | | 2 | 2 | 2 |

Tabella 8. L2 GR_COSTO

In L1, dato che la media del numero dei giovani ricercatori è identica, si nota che la media del costo dei GRF (giovani ricercatori con progetti diretti da donne) è di 239.809 nettamente inferiore rispetto a quella dei GRM (giovani ricercatori con progetti diretti da uomini) pari al 262.428 (“v.tab.7”).

Anche per L2 la media del numero dei GR (giovani ricercatori in generale) è molto simile, ma nonostante ciò le medie del costo dei GR sono differenti. Si vede che il costo GRF risulta 213.439, mentre quello riguardate i GRM è di 237.943. Anche in questa linea di ricerca i GRM hanno un costo più alto in media.

Se alle tab.7 e 8 si applica un operazione *roll-up* sulle PT per entrambe le linee di ricerca, con la variabile la TIPO_ENTE, si ottiene una PT bidimensionale con i soliti dati interni.

| | | SESSO | | |
|--------------------------------|------------------|----------------|----------------|------------------|
| TIPO_ENTE | Dati | F | M | Totale Risultato |
| Altro | Media - GR_COSTO | 238.523 | 255.461 | 248.936 |
| | Media - GR_NUM | 1,8 | 1,9 | 1,9 |
| CNR | Media - GR_COSTO | 280.091 | 284.298 | 282.222 |
| | Media - GR_NUM | 2,1 | 2,1 | 2,2 |
| Univ | Media - GR_COSTO | 236.058 | 260.815 | 247.853 |
| | Media - GR_NUM | 2 | 2 | 2 |
| Totale Media - GR_COSTO | | 239.809 | 262.438 | 250.860 |
| Totale Media - GR_NUM | | 2 | 2 | 2 |

Tabella 9. L1 GR_COSTO con operazione *roll-up*

- Il settore Altro ha una media di numero GR leggermente diversa fra

appartenenti a progetti diretti da uomini e donne. Comunque la media dei costi GR risulta a favore dei ricercatori referenti ai progetti coordinati da uomini.

- Il nel CNR ha il numero di GR è più alto rispetto a quello degli altri enti e risulta quasi identica per GRF e GRM. Si deduce che la media del numero di GR è più alta, conseguentemente anche il costo dei GR al CNR risulta più alto.
- Nel settore Universitario le medie del numero GR sono uguali. Differentemente la media del costo GRM risulta maggiore.

La media del costo GR è maggiore per i referenti ai capi progetto uomini, anche nel caso in cui la media del numero GR sia uguale ai progetti gestiti dal genere femminile.

Si guardino i risultati per L2:

| | | SESSO | | | |
|--------------------------------|------------------|-------|----------------|------------------|----------------|
| TIPO_ENTE | Dati | F | M | Totale Risultato | |
| Altro | Media - GR_COSTO | | 273.196 | 257.326 | 262.028 |
| | Media - GR_NUM | | 2,5 | 2,2 | 2,3 |
| CNR | Media - GR_COSTO | | 232.151 | 245.854 | 239.166 |
| | Media - GR_NUM | | 2,1 | 2,1 | 2,1 |
| Univ | Media - GR_COSTO | | 209.112 | 236.245 | 225.530 |
| | Media - GR_NUM | | 2,0 | 2,2 | 2,1 |
| Totale Media - GR_COSTO | | | 213.439 | 237.944 | 228.172 |
| Totale Media - GR_NUM | | | 2,0 | 2,2 | 2,1 |

Tabella 10. L2 GR_COSTO con operazione *roll-up*

- Il settore Altro ha una media di numero GRF maggiore rispetto rispetto agli altri. In questo caso la media dei costi GRF risulta maggiore rispetto a quella dei progetti diretti da uomini. Quindi i GRF hanno una media dei costi maggiore poiché hanno una media numero GR maggiore.
- Nel CNR la media numero GRF è vicina a quella GRM. Comunque la media costo GRM risulta maggiore rispetto a quella GRF.
- Nell'Università la media numero GRF è vicina a quella GRM.. Comunque la media costo GRM risulta maggiore rispetto a quella GRF.

La media del costo GR è maggiore per i referenti ai capi progetto uomini, anche nel caso in cui la media del numero GR sia inferiore a progetti gestiti dall'altro genere.

Quindi confrontando L1 e L2 per GR_COSTO si evidenzia che i GRM hanno mediamente un costo più elevato rispetto ai GRF.

Nel caso di **RCF_COSTO** (ricercatori di fama internazionale) si crea una PT con dimensione SESSO e con fatto la media di RCF_COSTO e la media di RCF_NUM.

| | SESSO | | |
|-------------------|-------|-------|------------------|
| Dati | F | M | Totale Risultato |
| Media - RCF_COSTO | 5.997 | 5.954 | 5.976 |
| Media - RCF_NUM | 0,2 | 0,2 | 0,2 |

Tabella 11. L1 RCF_COSTO

| | SESSO | | |
|-------------------|--------|--------|------------------|
| Dati | F | M | Totale Risultato |
| Media - RCF_COSTO | 12.600 | 12.922 | 12.794 |
| Media - RCF_NUM | 0,3 | 0,3 | 0,3 |

Tabella 12. L2 RCF_COSTO

Per L1, dato che le medie del numero dei RF (ricercatori di fama internazionale) sono identiche, si nota che la media del costo dei RF_F (ricercatori di fama internazionale referenti ai progetti diretti da donne) è di 5.997 è poco superiore a quella dei RF_M pari al 5.954 (“v.tab.11”).

Per L2 la media numero RF_M risulta leggermente maggiore di quella RF_F (“v.tab.12”).

Confrontando tab. 11 e tab. 12, L2 presenta costi RF in media più alti rispetto ad L1, perché L2 ha una media maggiore di numero partecipanti al progetto rispetto a L1, dato che più alta è la media del numero di RF più sarà elevata la media costo RF.

Se a quest'ultime tabelle si applica un *roll-up* sulle PT per entrambe le linee di ricerca, con la variabile la TIPO_ENTE, si ottiene una PT bidimensionale con i soliti dati interni.

| | | SESSO | | |
|---------------------------------|-------------------|--------------|--------------|------------------|
| TIPO_ENTE | Dati | F | M | Totale Risultato |
| Altro | Media - RCF_COSTO | 5.362 | 6.579 | 6.110 |
| | Media - RCF_NUM | 0,2 | 0,2 | 0,2 |
| CNR | Media - RCF_COSTO | 7.288 | 1.464 | 4.338 |
| | Media - RCF_NUM | 0,2 | 0,1 | 0,1 |
| Univ | Media - RCF_COSTO | 5.912 | 6.371 | 6.131 |
| | Media - RCF_NUM | 0,1 | 0,2 | 0,2 |
| Totale Media - RCF_COSTO | | 5.997 | 5.954 | 5.976 |
| Totale Media - RCF_NUM | | 0,2 | 0,2 | 0,2 |

Tabella 13. L1 RCF_COSTO con operazione *roll-up*

- Il settore Altro ha una media di numero RF_M uguale a RF_F. La media costo di RF_M è maggiore rispetto a quella di RF_F.
- Nel CNR la media numero RF_F più alta di quella RF_M. Mentre il costo medio di RF_F è più alto rispetto a quello di RF_M come si denota dalla PT.
- Nell' Università la media numero RF_M è lievemente maggiore rispetto a RC_F. Quindi la media costo RF_M è maggiore rispetto a quella RC_F.

La media del costo medio RF_M risulta maggiore rispetto a RF_F.

Si guardino i risultati per L2 :

| | | SESSO | | |
|---------------------------------|-------------------|---------------|---------------|------------------|
| TIPO_ENTE | Dati | F | M | Totale Risultato |
| Altro | Media - RCF_COSTO | 6.125 | 11.846 | 10.151 |
| | Media - RCF_NUM | 0,2 | 0,2 | 0,2 |
| CNR | Media - RCF_COSTO | 3.050 | 14.561 | 8.942 |
| | Media - RCF_NUM | 0,1 | 0,3 | 0,2 |
| Univ | Media - RCF_COSTO | 13.968 | 12.847 | 13.290 |
| | Media - RCF_NUM | 0,3 | 0,3 | 0,3 |
| Totale Media - RCF_COSTO | | 12.600 | 12.922 | 12.794 |
| Totale Media - RCF_NUM | | 0,3 | 0,3 | 0,3 |

Tabella 14. L2 RCF_COSTO con operazione *roll-up*

- Il settore Altro ha una media di numero RF_M è identica a RF_F. La media costo di RF_M è molto maggiore rispetto a quella di RF_F.
- Nel CNR la media numero RF_M è molto maggiore a quella RF_F. Quindi il costo medio di RF_M è molto maggiore rispetto a quello di RF_F come si denota dalla PT.
- Nell' Università le medie numero RF_M e RC_F sono identiche. Invece la

media costi RF_F è maggiore a RF_M.

La media del costo medio di RF_M risulta maggiore rispetto a RF_F. Si nota che al variare del numero medio di RF varia anche il costo medio degli RF. Però in alcuni casi se il numero medio di RF è pari per entrambi i generi, si nota che il costo medio degli RF è maggiore per i RF_M.

Paragonando le tab. 13 e 14 si nota che i costi medi di L2 sono nettamente superiori rispetto a quelli di L1.

Nel caso della variabile **COSTO_TOTALE** si crea una PT con dimensione SESSO e con dati la media di COSTO_TOTALE .

| SESSO | | |
|---------|---------|------------------|
| F | M | Totale Risultato |
| 885.809 | 937.951 | 911.273 |

Tabella 15. L1 COSTO TOTALE

| SESSO | | |
|---------|-----------|------------------|
| F | M | Totale Risultato |
| 980.946 | 1.056.242 | 1.026.215 |

Tabella 16. L2 COSTO TOTALE

Per L1 il costo medio di CT_M (costo progetti diretti da uomini) risulta maggiore rispetto a quello CT_F (progetti coordinati da donne) (“v.tab.15”).

Per L2 anche in questo caso il costo medio sul CT_M è maggiore di CT_F (“v.tab.16”).

Il costo medio dei progetti di L2 è nettamente superiore rispetto ad L1.

Se a quest'ultime tabelle si applica un *roll-up* sulle PT per entrambe le linee di ricerca, con la variabile la TIPO_ENTE, si ottiene una PT bidimensionale con i soliti dati interni.

| Media - COSTO_TOT | SESSO | | |
|-------------------------|----------------|----------------|------------------|
| TIPO_ENTE | F | M | Totale Risultato |
| Altro | 873.120 | 974.398 | 935.381 |
| CNR | 1.027.268 | 1.050.658 | 1.039.115 |
| Univ | 873.116 | 922.121 | 896.464 |
| Totale Risultato | 885.809 | 937.951 | 911.273 |

Tabella 17. L1 COSTO TOTALE con operazione *roll-up*

- Nel CNR si ha un costo medio dei progetti per CT_M maggiore rispetto a quello delle CT_F.
- Nel settore Universitario e Altro si ha ancora un costo più alto nel settore CT_M.

Nel complesso si nota che i progetti mediamente più costosi fanno riferimento al CNR, seguiti da Università ed Altro. Inoltre i progetti di CT_M sono mediamente più costosi di quelli CT_F.

| Media - COSTO_TOT | SESSO | | |
|-------------------------|----------------|------------------|------------------|
| TIPO_ENTE | F | M | Totale Risultato |
| Altro | 1.192.524 | 1.130.430 | 1.148.828 |
| CNR | 1.089.556 | 1.086.040 | 1.087.756 |
| Univ | 960.561 | 1.049.781 | 1.014.545 |
| Totale Risultato | 980.946 | 1.056.242 | 1.026.215 |

Tabella 18. L2 COSTO TOTALE con operazione *roll-up*

- Nel CNR si ha un costo medio dei progetti per CT_M leggermente maggiore rispetto a quello delle CT_F.
- Nell'Università si ha ancora un costo più alto nel settore CT_M.
- Nel settore Altro si nota un costo medio maggiore per CT_F rispetto a CT_M.

Nel complesso si evidenzia che i progetti mediamente più costosi fanno riferimento ad Altro, seguiti da CNR e Università. Inoltre i progetti di CT_M sono mediamente più costosi di quelli CT_F.

Confrontando i risultati delle tab. 17 e 18 si evince un maggiore costo medio per L2. Inoltre sia per le due linee di ricerca e sia per le diverse granularità, si delinea un costo medio maggiore per CT_M.

3.3 Discrimination Discovery

In questo paragrafo si illustrerà il concetto di scoperta della discriminazione, analizzando lo *step1* del processo di erogazione fondi sui progetti, nonché l'*international peer-review*. Successivamente si analizzerà lo *step2* del processo di selezione.

La fase di *international peer-review* sarà analizzata e discussa in 3 diverse fasi di analisi, dove le ultime due fasi saranno una conseguenza della prima:

1. Allo *step1* si esaminerà la percentuale di progetti ammessi e non su diverse granularità, utilizzando le PT.
2. Si analizzeranno i costi per valutare il costo medio dei progetti idonei e non idonei, controllando lo status dei progetti secondo la produttività. Le valutazioni di un progetto hanno un ruolo principale nel superamento dello *step 1*, poiché un progetto ha diritto ad essere ammesso alla fase successiva solo se ottiene il massimo su tutte e quattro le valutazioni. Studieremo quest'ultime grazie alle distribuzioni di frequenza e relativi istogrammi e se ne apprezzeranno e discuteranno le proprietà statistiche grazie agli indicatori sintetici per centralità e variabilità. Questa seconda fase di analisi viene eseguita per capire se la selezione che è stata fatta, è stata eseguita con criterio.
3. Si prenderanno in esame le quattro valutazioni che sono state ottenute da ogni progetto.

Mentre la fase *audit* sarà analizzata e discussa in due diverse fasi:

4. Si analizzerà il secondo e ultimo *step2* o *audit*, tramite le PT. È l'ultima fase prima di ricevere i finanziamenti sui progetti di ricerca eseguiti.
5. Si analizzeranno i costi per valutare il costo medio dei progetti finanziati e non, controllando lo status dei progetti secondo la produttività.

Infine si utilizzerà l'analisi di regressione per:

6. Studiare il rapporto fra le variabili tramite l'analisi di regressione e per comprendere da cosa derivano eventuali differenze fra progetti finanziati e non in base al sesso.

3.3.1 Analisi dell'*international peer-review*

In questa prima fase di analisi, per stimare in proporzione il numero di ammessi e non, si crea un PT con dimensione SESSO per AMMESSO e SESSO come dato da contare in percentuale.

| ContaNumeri - | SESSO | | |
|-------------------------|--------------|--------------|------------------|
| AMMESSO | F | M | Totale Risultato |
| no | 49,4% | 46,5% | 95,9% |
| yes | 1,7% | 2,4% | 4,1% |
| Totale Risultato | 51,2% | 48,8% | 100,0% |

Tabella 19. L1 percentuale ammessi/non ammessi

| ContaNumeri - | SESSO | | |
|-------------------------|--------------|--------------|------------------|
| AMMESSO | F | M | Totale Risultato |
| no | 38,3% | 55,1% | 93,4% |
| yes | 1,6% | 5,0% | 6,6% |
| Totale Risultato | 39,9% | 60,1% | 100,0% |

Tabella 20. L2 percentuale ammessi/non ammessi

Per le donne si registra una percentuale minore di accettazione di progetti, rispetto agli uomini:

- In L1 per le donne registriamo una percentuale minore di accettazione pari all'1,7 %.
- In L2 l' 1,56 % delle donne è ammessa allo *step2*.

Successivamente aumentando l'aggregazione dei dati tramite la *variabile TIPO_ENTE*, si otterrà un cubo bidimensionale SESSO x AMMESSO x TIPO_ENTE e SESSO come dato di conteggio in percentuale.

| ContaNumeri - SESSO | | SESSO | | |
|-------------------------|-----------|--------------|--------------|------------------|
| AMMESSO | TIPO_ENTE | F | M | Totale Risultato |
| no | Altro | 2,5% | 3,8% | 6,3% |
| | CNR | 4,1% | 4,0% | 8,1% |
| | Univ | 42,8% | 38,6% | 81,4% |
| yes | Altro | 0,1% | 0,3% | 0,4% |
| | CNR | 0,1% | 0,3% | 0,4% |
| | Univ | 1,5% | 1,8% | 3,3% |
| Totale Risultato | | 51,2% | 48,8% | 100,0% |

Tabella 21. L1 con *variabile TIPO_ENTE*

- L'ente con la maggior percentuale di ammessi risulta essere l'Università seguita da Altro e CNR.
- I capi progetto uomini hanno lo 0,3 % di ammessi per Altro a differenza delle donne pari solo allo 0,1 %.
- Per quanto concerne il CNR troviamo in vantaggio sempre gli uomini con lo 0,3 % contro lo 0,1 % delle donne.
- Nel ramo universitario abbiamo l'1,8 % di uomini ammessi rispetto all'1,5 % delle donne.

I settori dove le donne hanno minor successo sono quelli dell'Università e del CNR.

Invece i risultati per L2 sono:

| ContaNumeri - SESSO | | SESSO | | |
|-------------------------|-----------|--------------|--------------|------------------|
| AMMESSO | TIPO_ENTE | F | M | Totale Risultato |
| no | Altro | 1,1% | 2,6% | 3,7% |
| | CNR | 4,0% | 3,7% | 7,7% |
| | Univ | 33,2% | 48,8% | 82,0% |
| yes | Altro | 0,1% | 0,3% | 0,4% |
| | CNR | 0,1% | 0,7% | 0,8% |
| | Univ | 1,4% | 4,1% | 5,4% |
| Totale Risultato | | 39,9% | 60,1% | 100,0% |

Tabella 22. L2 con *variabile TIPO_ENTE*

- L'ente con maggior idonei risulta essere l'Università seguita da Altro e CNR.

- L'ente Altro risulta avere lo 0,3 % di capi progetto uomini ammessi, mentre solo lo 0,1 % per le donne.
- Nel fascia dedicata al CNR si riscontra che lo 0,7 % dei progetti coordinati da uomini ha avuto un risultato positivo in questa prima fase, differentemente le donne hanno ottenuto lo 0,1 %.
- Nel caso dell'Università ci sono percentuali di ammessi più alte: il 4,1 % progetti diretti da uomini e l'1,4 % da donne.

Nella linea di ricerca L2 l'organo con maggior ammessi è sempre l'Università, dove si nota un grande divario fra progetti diretti da uomini e donne. Quest'ultima caratteristica la si scorge anche nella linea di ricerca dei giovani, ma non risulta così evidente come in questo caso.

Continuando l'ispezione di ammessi e non più nello specifico, si aggiunge ancora un grado di aggregazione grazie alla *variabile SETT_ID_1_LV_1* che delinea le aree di ricerca e studio. Si ottiene così un cubo bidimensionale SESSO x (AMMESSO x TIPO_ENTE x SETT_ID_1_LV_1) e SESSO come dato di conteggio in percentuale. Questa variabile assume valori rispetto alle tre categorie:

- PE (*Physical sciences and Engineering*)
- SH (*Social Sciences and Humanities*)
- LC (*Life Sciences*)

| ContaNumeri - SESSO | | | SESSO | | |
|---------------------|--------------|---------|-------|-------|------------------|
| TIPO_ENTE | SETT_ID_1_LV | AMMESSO | F | M | Totale Risultato |
| Altro | LS | no | 0,6% | 0,7% | 1,3% |
| | | yes | 0,1% | | 0,1% |
| | PE | no | 1,4% | 2,5% | 4,0% |
| | | yes | | 0,2% | 0,2% |
| | SH | no | 0,4% | 0,6% | 1,0% |
| | | yes | | 0,2% | 0,2% |
| CNR | LS | no | 1,7% | 0,7% | 2,4% |
| | | yes | 0,1% | 0,1% | 0,2% |
| | PE | no | 2,1% | 3,1% | 5,2% |
| | | yes | | 0,1% | 0,1% |
| | SH | no | 0,4% | 0,2% | 0,6% |
| | | yes | 0,1% | 0,1% | 0,1% |
| Univ | LS | no | 17,7% | 8,9% | 26,6% |
| | | yes | 0,7% | 0,5% | 1,2% |
| | PE | no | 12,0% | 19,6% | 31,6% |
| | | yes | 0,4% | 0,7% | 1,1% |
| | SH | no | 13,2% | 10,1% | 23,3% |
| | | yes | 0,4% | 0,6% | 1,0% |

Tabella 23. L1 con variabile *SETT_ID_1_LV_1*

- L'ente Altro mostra che per il settore LS solo lo 0,1 % dei progetti diretti da donne ha l'idoneità per lo *step* successivo. Mentre in PE e SH la situazione si ribalta a favore degli uomini, dei quali lo 0,2 % è ammesso alla fase seguente. Quindi si nota che i progetti diretti da donne si affermano solo nel campo di ricerca LS e gli uomini nei restanti campi di studio.
- Il CNR delinea per LS che lo 0,6 % di progetti coadiuvati da donne e lo 0,1 % diretto da uomini sono idonei. Mentre per PE lo 0,1 % degli uomini è ammesso. Per SH lo 0,6 % è ammesso per i progetti coordinati da entrambi i sessi. Si nota che i progetti diretti da donne si affermano solo nei settori LS e SH, invece gli uomini si confermano su tutti e tre i campi di studio e con percentuali più alte.
- L'Università evidenzia lo 0,1 % di ammesse per i progetti di ricerca e lo 0,5 % di ammessi per il campo di studio LS. Mentre per PE uno 0,7 % va a gli uomini e lo 0,4 % alle donne. Infine il settore SH mostra che lo 0,4 % di progetti diretti da donne e lo 0,6 % progetti diretti da uomini sono ammessi.

I progetti dei settori PE (*physical sciences and engineering*) e SH (*social sciences and Humanities*) registrano un minor successo oppure pari allo 0 % per i progetti diretti da donne. Questa caratteristica è meno evidente nell'università, dove l'affermazione dei progetti coordinati da donne risulta migliore.

- I progetti diretti da donne ottengono maggiori rifiuti nel campo PE e SH, mentre si esaltano in LS.
- Gli uomini raccolgono meno idoneità nel campo di studio LS, diversamente nei settori PH e SH risultano avere una percentuale di idonei più alta.

| ContaNumeri - SESSO | | | SESSO | | Totale Risultato |
|---------------------|--------------|---------|-------|-------|------------------|
| TIPO_ENTE | SETT_ID_1_LV | AMMESSO | F | M | |
| Altro | LS | no | 0,2% | 0,4% | 0,6% |
| | | yes | 0,1% | 0,3% | 0,4% |
| | PE | no | 0,9% | 2,1% | 3,0% |
| | | yes | 0,1% | 0,3% | 0,4% |
| SH | no | | | 0,1% | 0,1% |
| | yes | | | | |
| CNR | LS | no | 1,5% | 0,6% | 2,0% |
| | | yes | 0,1% | 0,1% | 0,2% |
| | PE | no | 2,4% | 3,1% | 5,4% |
| | | yes | 0,1% | 0,5% | 0,6% |
| | SH | no | 0,2% | 0,1% | 0,3% |
| | | yes | | | 0,1% |
| Univ | LS | no | 12,5% | 12,1% | 24,6% |
| | | yes | 0,8% | 1,1% | 1,8% |
| | PE | no | 10,4% | 24,4% | 34,8% |
| | | yes | 0,2% | 2,0% | 2,1% |
| | SH | no | 10,3% | 12,4% | 22,7% |
| | | yes | 0,5% | 1,1% | 1,5% |

Tabella 24. L2 con variabile *SETT_ID_1_LV_1*

- L'ente Altro mostra che per il settore LS e SH nessuno ha l'idoneità per lo *step* successivo. Mentre in PE lo 0,3 % dei progetti diretti da uomini e lo 0,1 % dei progetti coordinati da donne è ammesso alla fase seguente. Si nota che solo il campo PE ottiene dei *feed-back* positivi da entrambi i sessi, in particolare per il sesso maschile.

- Il CNR delinea per LS che lo 0,1 % di progetti diretti da donne e lo 0,1 % diretto da uomini sono idonei. Mentre per PE lo 0,5 % dei progetti coordinati da uomini e lo 0,05 % dei progetti diretti da donne è ammesso. Per SH lo 0,1 % è ammesso per i progetti coordinati da uomini. Si denota una importante plus-valenza a favore degli uomini in tutti e tre i campi di studio.
- L'Università evidenzia lo 0,8 % coordinati da donne per i progetti di ricerca e l'1,1 % progetti condotti da uomini per il campo di studio LS. Mentre per PE l'2 % va a progetti gestiti da uomini e lo 0,2 % alle donne. Infine il settore SH mostra che lo 0,5 % di progetti diretti da donne e l'1,1 % progetti diretti da uomini sono ammessi.

Confrontando tab. 23 e tab. 24, si evince che:

- Nei progetti coordinati da donne si nota una percentuale minore o pari allo 0 % per il campo di ricerca SH in tutti e tre gli organi, rispetto ai colleghi maschi.
- Nel campo di studio PE si nota una grande differenza in questo caso fra progetti diretti da uomini e donne per idonei e non, per entrambe le linee di ricerca.
- Per quanto riguarda il campo LS si nota una minor differenza rispetto ai precedenti.

Quindi:

- In entrambe le linee di ricerca la maggior parte dei capi progetto fa riferimento all'Università.
- Si evince già un certa divario fra i due sessi in base ai progetti ammessi e non.
- Si capisce che i progetti condotti da donne in particolare nei settori SH e PE ottengono molti meno consensi rispetto a quelli dei colleghi maschi.

Nel prossimo punto si valuterà in base all'analisi dei costi e produttività, se queste differenze continuano tra i progetti condotti da uomini e donne.

3.3.2 Analisi di costi e produttività *step 1*

Nel sottoparagrafo precedente si è valutato e passato in rassegna i risultati ottenuti per l'*internation peer-review*, analizzando questa fase secondo diverse variabili. I risultati ottenuti mostrano una certa disparità fra i progetti coordinati da uomini e donne. Questo paragrafo si divide in due fasi:

- a) Analisi sui costi dei progetti ammessi e non al fine di valutare una possibile disparità.
- b) Analisi della produttività in base al numero di pubblicazioni e alla media degli autori.

Nella **FASE A** si crea un PT con dimensione SESSO x AMMESSO e con dati la media dei costi totali.

| Media - COSTO_TOT | AMMESSO | | |
|-------------------------|----------------|----------------|------------------|
| SESSO | no | yes | Totale Risultato |
| F | 887.630 | 833.431 | 885.809 |
| M | 933.124 | 1.032.031 | 937.951 |
| Totale Risultato | 909.667 | 948.834 | 911.273 |

Tabella 25. L1 Fase A

| Media - COSTO_TOT | AMMESSO | | |
|-------------------------|------------------|------------------|------------------|
| SESSO | no | yes | Totale Risultato |
| F | 971.105 | 1.222.528 | 980.946 |
| M | 1.057.941 | 1.037.653 | 1.056.242 |
| Totale Risultato | 1.022.317 | 1.081.402 | 1.026.215 |

Tabella 26. L2 Fase A

Nella tab. 25 si nota che la media di M_YES (progetti ammessi con capi progetto uomini) è nettamente superiore a quella di F_YES (progetti ammessi con capi progetto donne) e precisamente la differenza fra le due medie è 198.599. Mentre la media di M_NO (progetti non idonei con capi progetto uomini) è molto più vicina alla media di F_NO (progetti non idonei con capi progetto donne), e la differenza fra le due medie è 45.494. Ergo che i progetti ammessi per M_YES hanno un costo medio più alto rispetto a quello di F_YES.

Nella tab. 26 la situazione è ribaltata rispetto alla tab. 25; la media di M_YES è nettamente inferiore a quella di F_YES e la differenza fra le due medie è 184.874. Mentre la media di M_NO è maggiore alla media di F_NO e la differenza fra le due medie è dello'86.836. Ergo che i progetti ammessi per M_YES hanno un costo medio più alto rispetto a quello di F_YES.

Confrontando L1 e L2 si nota che: in L1 si nota un maggiore favoritismo di M_YES, mentre in L2 la differenza fra F_YES e M_YES risulta minore. In sintesi M_YES hanno un costo maggiore rispetto a F_YES.

Nella **FASE B** si crea un PT con dimensione SESSO x NUM_PUBB e con dati la media NUM_PUBB.

| Media - NUM_PUBB | SESSO | | |
|-------------------------|-------------|-------------|------------------|
| AMMESSO | F | M | Totale Risultato |
| no | 10,5 | 12,3 | 11,4 |
| yes | 11,7 | 13,7 | 12,9 |
| Totale Risultato | 10,6 | 12,3 | 11,4 |

Tabella 27. L1 Fase B

| Media - NUM_PUBB | SESSO | | |
|-------------------------|-------------|-------------|------------------|
| AMMESSO | F | M | Totale Risultato |
| no | 19,6 | 22,1 | 21,1 |
| yes | 18,3 | 21,2 | 20,5 |
| Totale Risultato | 19,5 | 22,1 | 21,1 |

Tabella 28. L2 Fase B

Si nota una maggiore produttività in L2 rispetto ad L1. Inoltre i progetti coordinati da uomini mostrano una maggiore produttività rispetto ai progetti diretti da donne in entrambe le linee di ricerca.

Nel prossimo punto si valuterà, in base all'analisi dei voti, il motivo di queste differenze fra progetti condotti da uomini e donne.

3.3.3 Distribuzioni dei voti

Nel sottoparagrafo precedente si sono valutati e passati in rassegna i risultati ottenuti dall'analisi dei costi sui progetti idonei e non, analizzando questa fase

secondo diverse variabili; successivamente si è analizzata la produttività dei progetti. I risultati ottenuti mostrano una certa disparità fra i progetti coordinati da uomini e donne. Adesso si valuteranno i risultati precedentemente ottenuti secondo la qualità. Questo passo è fondamentale per comprendere se le esclusioni eseguite da parte della commissione allo *step1* sono discriminatorie o meno. Si analizzerà la qualità dei progetti tramite i quattro criteri di valutazione:

- *Scientific relevance (score 0-8)*: voto in ottavi, misura la rilevanza scientifica
- *Impact of the project (score 0-7)*: voto in settimi, rileva che impatto ha il progetto sulla commissione
- *Scientific and technical value (score 0-15)*: voto in quindicesimi
- *Quality of the partnership (score 0-10)*: voto in decimi, misura la qualità della collaborazione.

Si analizzerà questo insieme dei dati con l'aiuto di diagrammi a barre rappresentate da distribuzioni di frequenza. Quest'ultime offrono una lettura rapida delle caratteristiche più importanti della serie di dati. La rappresentazione grafica consente una visualizzazione immediata dell'andamento del fenomeno e della struttura della distribuzione di frequenze e quindi una più efficace e globale descrizione dei dati. Inoltre, permette di confrontare agevolmente più distribuzioni e di individuare peculiarità dei dati e in questo caso in particolare le differenze fra capi progetto uomini e donne. Rispetto alle cifre, le figure forniscono impressioni che sono percepite con maggiore facilità, ma nel contempo hanno il limite di essere meno precise e meno ricche di particolari. Per i caratteri qualitativi, la tabella e le rappresentazioni grafiche esauriscono quasi completamente gli aspetti descrittivi, quando sia possibile leggere con esattezza le frequenze delle varie classi. Per i caratteri quantitativi, si pone il problema di sintesi oggettive che possano essere elaborate matematicamente e quindi che siano numeriche, al fine di un'analisi obiettiva. Si descriverà l'analisi compiuta sui quattro criteri di valutazione secondo tre proprietà principali:

1. Indicatori di centralità: servono per individuare il valore intorno al quale i dati sono raggruppati; le misure utilizzate sono essenzialmente tre:
 - Moda
 - Media
 - Mediana

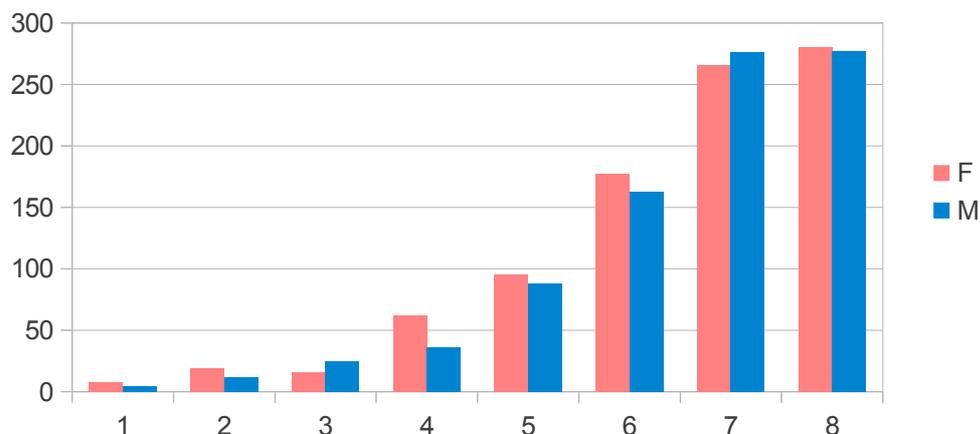
2. Dispersione o variabilità: è la seconda importante caratteristica di una distribuzione di dati, definisce la forma più o meno raccolta della distribuzione intorno al valore centrale e fornisce indicazioni sul tipo di test da applicare. Le misure utilizzate sono essenzialmente due:
 - Varianza
 - Deviazione standard

3. Forma: le analisi della forma di una distribuzione statistica sono limitate alla semplice descrizione della forma della distribuzione e riguardano due caratteristiche:
 - Asimmetria
 - Curtosi

I risultati ottenuti dalle distribuzioni di frequenza eseguite per le quattro variabili di valutazione per la variabile qualitativa SESSO, mostrano per **L1**:

1. *Scientific Relevance*

Distribuzione di frequenza Scientific Relevance (score 0-8)



| INDICATORI DI CENTRALITÀ | | F | M |
|-------------------------------------|--|------|------|
| MODA | | 8 | 8 |
| MEDIANA | | 7 | 7 |
| MEDIA | | 6,5 | 6,6 |
| INDICATORI DI VARIABILITÀ | | F | M |
| VARIANZA | | 2,3 | 2,1 |
| DEVIAZIONE STANDARD | | 1,5 | 1,4 |
| MISURE DI RAGGRUPPAMENTO | | F | M |
| Q1 | | 6 | 6 |
| Q2 | | 7 | 7 |
| Q3 | | 8 | 8 |
| Q4 | | 8 | 8 |
| MISURE DI DISPERSIONE E VARIABILITÀ | | F | M |
| ASIMMETRIA | | -1,2 | -1,3 |
| CURTOSI | | 1,3 | 1,5 |

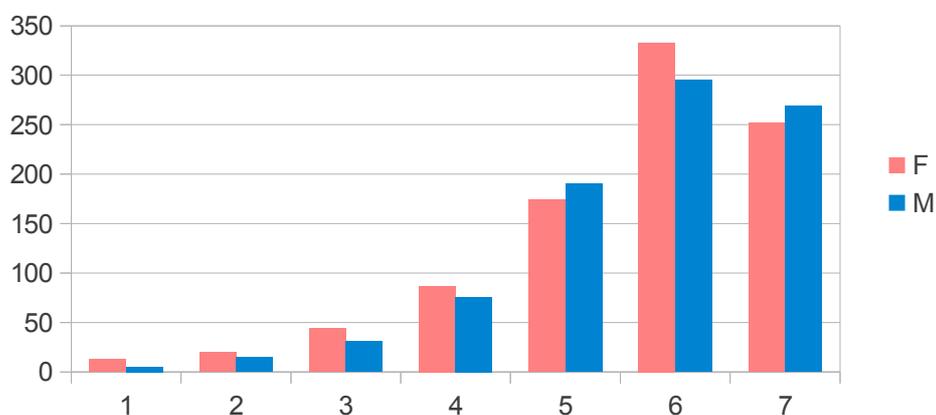
Tabella 29. L1 distribuzione di frequenza *Scientific Relevance*

- Per entrambi i sessi si evidenzia un evidente andamento asimmetrico, con coda a sinistra. Tale circostanza è confermata sia dall'indice di asimmetria negativo, che dagli indicatori di centralità media, mediana, moda.
- La media dei progetti coordinati da uomini risulta lievemente maggiore rispetto a quella delle donne. Essendo positivo l'indice di curtosi definisce una curva *leptocurtica*, ossia più "appuntita" di una curva normale.
- Gli indicatori di variabilità dimostrano una maggiore dispersione dei dati nelle distribuzione delle donne rispetto a quella degli uomini.
- Nonostante ciò, la variabilità basata sui centili dimostra una differenza

piccola fra Q3 e Q1, segnale che i dati sono raggruppati.

2. *Impact of the Project*

Distribuzione di frequenza Impact of project (score 0-7)



| INDICATORI DI CENTRALITÀ | | F | M |
|----------------------------|--|------|------|
| MODA | | | 6 |
| MEDIANA | | 6 | 6 |
| MEDIA | | 5,6 | 5,7 |
| INDICATORI DI VARIABILITÀ | | F | M |
| VARIANZA | | 1,8 | 1,5 |
| DEVIAZIONE STANDARD | | 1,3 | 1,2 |
| MISURE DI RAGGRUPPAMENTO | | F | M |
| Q1 | | 5 | 5 |
| Q2 | | 6 | 6 |
| Q3 | | 7 | 7 |
| Q4 | | 7 | 7 |
| DISTRIBUZIONI DI FREQUENZE | | F | M |
| ASIMMETRIA | | -1,2 | -1,1 |
| CURTOSI | | 1,3 | 1,2 |

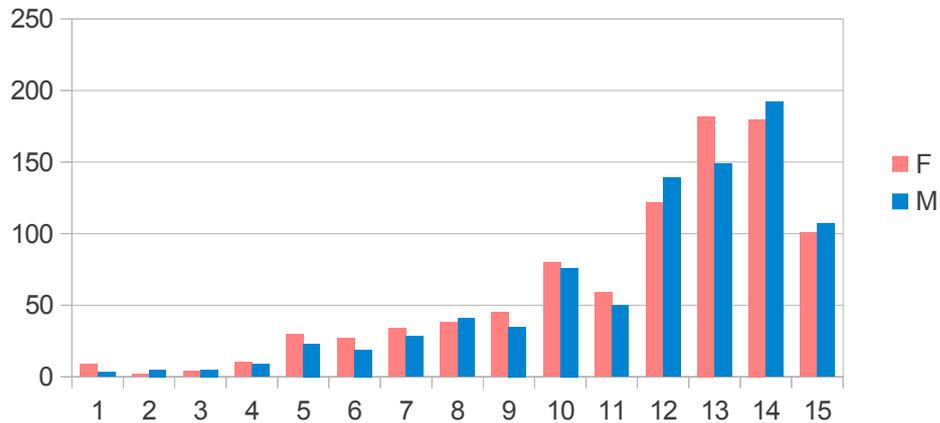
Tabella 30. L1 distribuzione di frequenza *Impact of the project*

- Per entrambi i sessi si denota un evidente andamento asimmetrico, con coda a sinistra. Tale circostanza è confermata sia dall'indice di asimmetria negativo, che dagli indicatori di centralità media, mediana, moda.
- La media risulta lievemente a favore degli uomini. Essendo positivo l'indice di curtosi definisce una curva *leptocurtica*, in questo caso è più appuntita la distribuzione dei progetti diretti da donne.
- Gli indicatori di variabilità dimostrano una maggiore dispersione dei dati nella distribuzione delle donne rispetto a quella degli uomini.

- Nonostante ciò le misure di raggruppamento, basate sui quartili, mostrano una differenza piccola fra Q3 e Q1, segnale che i dati sono raggruppati.

3. *Scientific and technical value*

Distribuzione di frequenza Scientific and technical value (score 0-15)



| INDICATORI DI CENTRALITÀ | | F | M |
|----------------------------|--|------|------|
| MODA | | | 13 |
| MEDIANA | | 13 | 13 |
| MEDIA | | 11,5 | 11,7 |
| INDICATORI DI VARIABILITÀ | | F | M |
| VARIANZA | | 9,3 | 8,5 |
| DEVIATIONE STANDARD | | 3,0 | 2,9 |
| MISURE DI RAGGRUPPAMENTO | | F | M |
| Q1 | | 10 | 10 |
| Q2 | | 13 | 13 |
| Q3 | | 14 | 14 |
| Q4 | | 15 | 15 |
| DISTRIBUZIONI DI FREQUENZE | | F | M |
| ASIMMETRIA | | -1,2 | -1,2 |
| CURTOSI | | 0,9 | 1,0 |
| ERRORE STANDARD | | 1,4 | 1,2 |

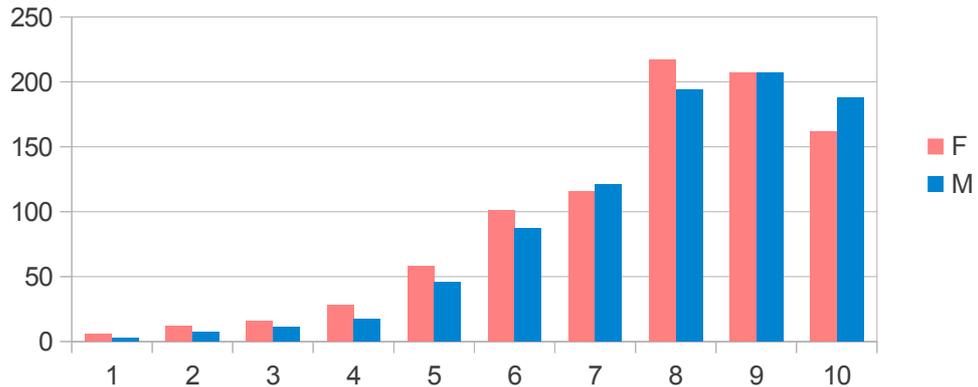
Tabella 31. L1 distribuzione di frequenza *Scientific and technical value*

- Per entrambi i sessi si ha un evidente andamento asimmetrico, con coda a sinistra. Tale circostanza è confermata dall'indice di asimmetria negativo.
- La media risulta leggermente a favore degli uomini. L'indice di curtosi svela che entrambe le distribuzioni sono leptocurtiche, in particolare nel caso dei progetti coordinati da uomini.
- Gli indicatori di variabilità dimostrano una maggiore dispersione dei dati nelle distribuzione delle donne rispetto a quella degli uomini.

- Nonostante ciò la variabilità basata sui quartili dimostra una differenza piccola fra Q3 e Q1, segnale che i dati sono raggruppati.

4. *Quality of the partnership*

Distribuzione di frequenza Quality of partnership (score 0-10)



| INDICATORI DI CENTRALITÀ | | F | M |
|-------------------------------------|--|------|------|
| MODA | | 8 | 9 |
| MEDIANA | | 8 | 8 |
| MEDIA | | 7,7 | 8,0 |
| INDICATORI DI VARIABILITÀ | | F | M |
| VARIANZA | | 3,7 | 8,5 |
| DEVIAZIONE STANDARD | | 1,5 | 1,8 |
| MISURE DI RAGGRUPPAMENTO | | F | M |
| Q1 | | 7 | 7 |
| Q2 | | 8 | 8 |
| Q3 | | 9 | 9 |
| Q4 | | 10 | 10 |
| MISURE DI DISPERSIONE E VARIABILITÀ | | F | M |
| ASIMMETRIA | | -1,0 | -1,0 |
| CURTOSI | | 0,8 | 1,0 |
| ERRORE STANDARD | | 0,5 | 0,6 |

Tabella 32. L1 distribuzione di frequenza *Quality of the partnership*

- Per entrambi i sessi si nota un evidente andamento asimmetrico, con coda a sinistra. Tale circostanza è confermata dall'indice di asimmetria negativo.
- La media risulta leggermente a favore degli uomini. L'indice di curtosi svela che entrambe le distribuzioni sono leptocurtiche, in particolare nel caso dei

progetti diretti da uomini.

- Gli indicatori di variabilità dimostrano una maggiore dispersione dei dati nelle distribuzioni delle donne rispetto a quella degli uomini.
- Nonostante ciò la variabilità basata sui quartili dimostra una differenza piccola nello scarto interquartile, segnale che i dati sono raggruppati. Dai quartili si nota una maggiore frequenza dei valori più alti.

Quindi riepilogando:

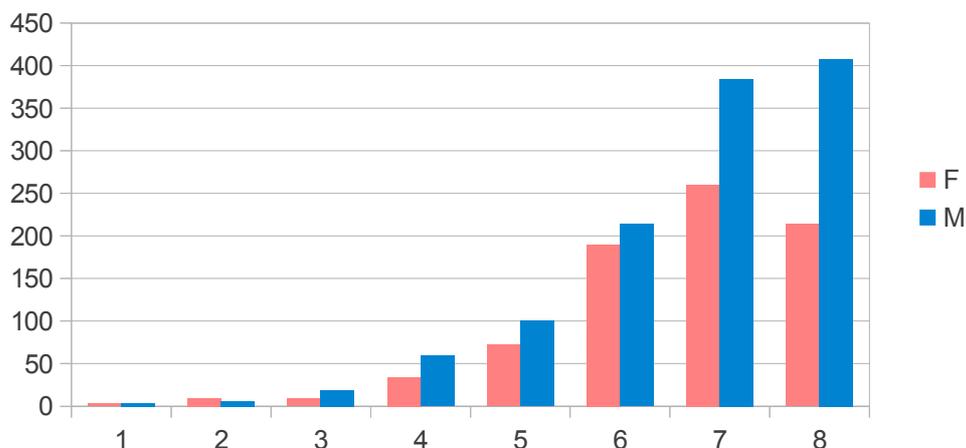
1. Gli indicatori di centralità non mostrano grandi disparità fra progetti diretti da uomini e donne. Le medie differiscono solo per lo 0,1 o lo 0,2 in favore dei progetti diretti da uomini.
2. Gli indici di variabilità dimostrano che i dati dei progetti coordinati da donne hanno leggermente una maggiore variabilità rispetto ai progetti diretti da uomini. Questo significa che i valori dei capi progetto donne si discostano poco di più dalla media, rispetto agli altri.
3. Gli indicatori di raggruppamento esprimono sempre le solite proprietà.
4. Le misure di dispersione e variabilità dimostrano che l'asimmetria delle distribuzioni maschili risultano leggermente più asimmetriche, rispetto a quelle dei capi progetto donne.

I non ammessi in **L1** sono potenzialmente discriminati, perché le distribuzioni non presentano differenze tali da attribuire una disparità come nel paragrafo precedente (“v. par. 3.3.1”). Inoltre notiamo che molto spesso nelle distribuzioni di capi progetto donne si raggiunge il picco massimo di frequenza nei valori adiacenti al massimo (“v. tab. 30,31 e 32”).

Si guardino i risultati per **L2**:

1. *Scientific Relevance*

Distribuzione di frequenza Scientific relevance (score 0-8)



| INDICATORI DI CENTRALITÀ | F | M |
|--------------------------|---|-----|
| MODA | | 7 |
| MEDIANA | | 7 |
| MEDIA | | 6,6 |

| INDICATORI DI VARIABILITÀ | F | M |
|---------------------------|-----|-----|
| VARIANZA | 1,7 | 1,7 |
| DEVIAZIONE STANDARD | 1,3 | 1,3 |

| MISURE DI RAGGRUPPAMENTO | F | M |
|--------------------------|---|---|
| Q1 | | 6 |
| Q2 | | 7 |
| Q3 | | 8 |
| Q4 | | 8 |

| DISTRIBUZIONI DI FREQUENZE | F | M |
|----------------------------|------|------|
| ASIMMETRIA | -1,2 | -1,2 |
| CURTOSI | 2,0 | 1,6 |

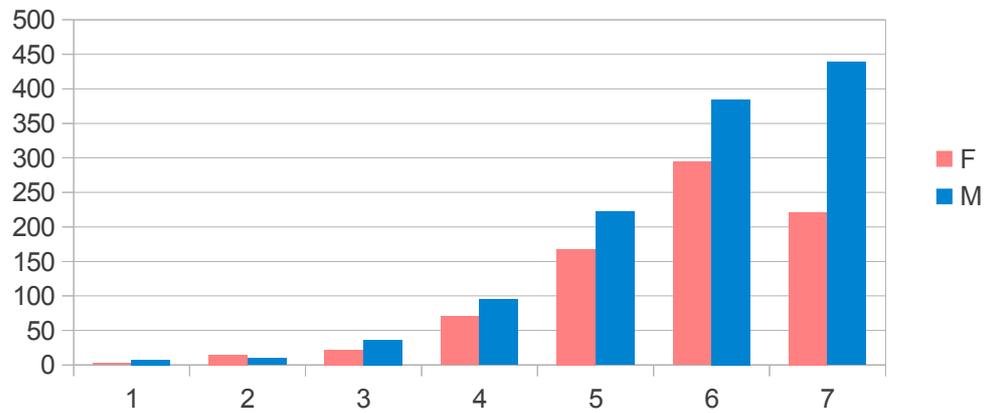
Tabella 33. L2 distribuzione di frequenza *Scientific Relevance*

- Per entrambi i sessi si evidenzia un evidente andamento asimmetrico, con coda a sinistra. Tale circostanza è confermata dall'indice di asimmetria negativo.
- La media risulta leggermente a favore degli uomini. L'indice di curtosi svela che entrambe le distribuzioni sono leptocurtiche, in particolare nel caso dei progetti diretti da donne.
- Gli indicatori di variabilità dimostrano una dispersione dei dati pressoché identiche per entrambi i sessi. Nonostante ciò la variabilità basata sui quartili mostra una differenza piccola fra Q3 e Q1, segnale che i dati sono

raggruppati.

2. *Impact of the project*

Distribuzione di frequenza Impact of the project (score 0-7)



| INDICATORI DI CENTRALITÀ | | F | M |
|--------------------------|--|-----|-----|
| MODA | | 6 | 7 |
| MEDIANA | | 6 | 6 |
| MEDIA | | 5,7 | 5,9 |

| INDICATORI DI VARIABILITÀ | | F | M |
|---------------------------|--|-----|-----|
| VARIANZA | | 1,4 | 1,4 |
| DEVIAZIONE STANDARD | | 1,2 | 1,2 |

| MISURE DI RAGGRUPPAMENTO | | F | M |
|--------------------------|--|---|---|
| Q1 | | 5 | 5 |
| Q2 | | 6 | 6 |
| Q3 | | 7 | 7 |
| Q4 | | 7 | 7 |

| DISTRIBUZIONI DI FREQUENZE | | F | M |
|----------------------------|--|------|------|
| ASIMMETRIA | | -1,1 | -1,2 |
| CURTOSI | | 1,2 | 1,5 |

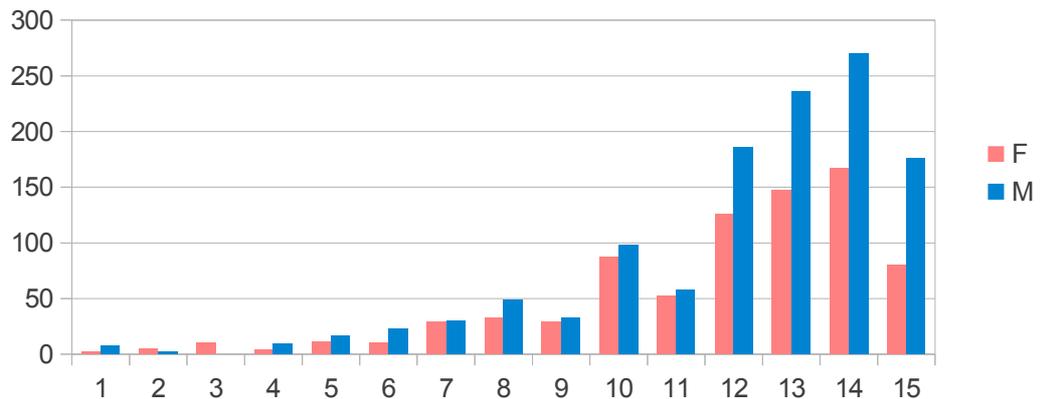
Tabella 34. L2 distribuzione di frequenza *Impact of the project*

- Per entrambi i sessi il grafico mostra un evidente andamento asimmetrico, con coda a sinistra. Tale circostanza è confermata sia dall'indice di asimmetria negativo, che dagli indicatori di centralità media, mediana, moda.
- Notiamo che la distribuzione dei progetti diretti da uomini ha una asimmetria maggiore rispetto a quelle delle donne.
- La media risulta leggermente a favore degli uomini. L'indice di curtosi svela che entrambe le distribuzioni sono leptocurtiche, in particolare nel caso dei progetti diretti da uomini.

- Gli indicatori di variabilità dimostrano una maggiore dispersione dei dati nei progetti diretti da uomini. Infine la variabilità basata sui quartili dimostra una differenza piccola fra Q3 e Q1, segnale che i dati sono raggruppati verso i valori medio alti per entrambi i sessi.

3. *Scientific and technical value*

Distribuzione di frequenza Scientific and technical value (score 0-15)



| INDICATORI DI CENTRALITÀ | | F | M |
|--------------------------|--|------|------|
| MODA | | | 14 |
| MEDIANA | | | 13 |
| MEDIA | | 11,7 | 12,1 |

| INDICATORI DI VARIABILITÀ | | F | M |
|---------------------------|--|-----|-----|
| VARIANZA | | 7,7 | 7,1 |
| DEVIAZIONE STANDARD | | 2,8 | 2,7 |

| MISURE DI RAGGRUPPAMENTO | | F | M |
|--------------------------|--|----|----|
| Q1 | | 10 | 11 |
| Q2 | | 12 | 13 |
| Q3 | | 14 | 14 |
| Q4 | | 15 | 15 |

| DISTRIBUZIONI DI FREQUENZE | | F | M |
|----------------------------|--|------|------|
| ASIMMETRIA | | -1,3 | -1,4 |
| CURTOSI | | 1,6 | 2,1 |

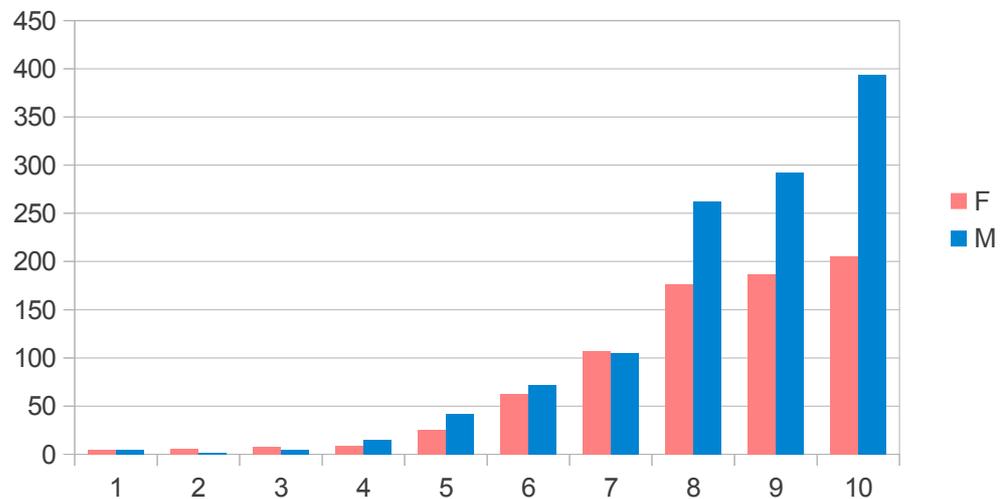
Tabella 35. L2 distribuzione di frequenza *Scientific and technical value*

- Si evidenzia per entrambi i sessi un chiaro andamento asimmetrico, con coda a sinistra. Tale circostanza è confermata sia dall'indice di asimmetria negativo, che dagli indicatori di centralità media, mediana, moda.

- Si nota che la distribuzione dei progetti diretti da uomini ha una asimmetria maggiore rispetto a quelle delle donne.
- La media risulta leggermente a favore degli uomini. L'indice di curtosi svela che entrambe le distribuzioni sono leptocurtiche, in particolare nel caso dei progetti diretti da uomini.
- Gli indicatori di variabilità dimostrano una maggiore dispersione dei dati nei progetti diretti da donne. Infine la variabilità basata sui quartili dimostra una differenza piccola fra Q3 e Q1, segnale che i dati sono raggruppati verso i valori medio alti per entrambi i sessi, anche se i valori dei progetti diretti da donne in questo caso sono spostati leggermente più in basso, poiché lo scarto interquartile ($Q3 - Q1$) risulta maggiore per le donne che per gli uomini.

4. *Quality of partnership*

Distribuzione di frequenza Quality of partnership (score 0-10)



| INDICATORI DI CENTRALITÀ | F | M |
|----------------------------|---|------|
| MODA | | 10 |
| MEDIANA | | 8 |
| MEDIA | | 8,2 |
| | | 10 |
| | | 9 |
| | | 8,5 |
| INDICATORI DI VARIABILITÀ | F | M |
| VARIANZA | | 3,0 |
| DEVIAZIONE STANDARD | | 1,7 |
| | | 2,6 |
| | | 1,6 |
| MISURE DI RAGGRUPPAMENTO | F | M |
| Q1 | | 7 |
| Q2 | | 8 |
| Q3 | | 10 |
| Q4 | | 10 |
| | | 8 |
| | | 9 |
| | | 10 |
| | | 10 |
| DISTRIBUZIONI DI FREQUENZE | F | M |
| ASIMMETRIA | | -1,3 |
| CURTOSI | | 2,2 |
| | | -1,4 |
| | | 2,3 |

Tabella 36. L2 distribuzione di frequenza *Quality of partnership*

- Il diagramma a barre evidenzia per entrambi i sessi un chiaro andamento asimmetrico, con coda a sinistra. Tale circostanza è confermata sia dall'indice di asimmetria negativo.
- Si nota che la distribuzione dei progetti diretti da uomini ha una asimmetria maggiore rispetto a quelle delle donne.
- La media risulta leggermente a favore degli uomini. L'indice di curtosi svela che entrambe le distribuzioni sono leptocurtiche, in particolare nel caso dei progetti coordinati da uomini.
- Gli indicatori di variabilità dimostrano una maggiore dispersione dei dati nei progetti diretti da donne. Infine la variabilità basata sui quartili dimostra una differenza piccola fra Q3 e Q1, segnale che i dati sono raggruppati verso i valori medio alti per entrambi i sessi, anche se i valori dei progetti diretti da donne in questo caso sono spostati leggermente più in basso, poiché lo scarto interquartile (Q3 - Q1) risulta maggiore per le donne che per gli uomini.

I risultati di queste distribuzioni confermano come potenzialmente discriminatori i risultati precedenti: c'è una forte discrepanza fra i progetti ammessi e non in base al sesso. Questo perché:

- non si notano grandi differenze nelle forme delle distribuzioni in base al genere.

- le medie dei voti differiscono solo per lo 0,2.
- le misure di raggruppamento evidenziano che i valori dei progetti coordinati da uomini e donne sono più raggruppati verso i valori alti.

Confrontando L1 e L2 per i risultati delle distribuzioni si nota che:

- I risultati dei diagrammi sulla linea di ricerca L1 mostrano una disparità fra ammessi e non in base al sesso dei coordinatori, perché tutte le distribuzioni delle donne hanno un picco nel valore massimo dell'ascissa. Mentre le distribuzioni degli uomini risultano più lineari.
- Il divario fra le due linee di ricerca, è dato dal fatto che le distribuzioni dei capi progetto donne raggiungono le frequenze massime nei punti dell'ascissa vicini al valore massimo. Questo comporta un maggior picco nel valore massimo. Mentre le distribuzioni maschili si dimostrano più lineari.
- In L2 le misure di raggruppamento svelano una maggiore asimmetria delle distribuzioni sui progetti diretti da uomini, poiché hanno uno scarto interquartile minore di quello dei progetti diretti da donne. Questo significa che i dati degli uomini sono maggiormente concentrati verso i valori alti del diagramma.

Si notano delle piccole differenze in base al genere, date dalle proprietà della statistica descrittiva per L1 e L2, che tuttavia non giustificano un così grande divario fra capi progetto uomini e donne per entrambe le linee di ricerca.

3.3.4 Analisi dell'*Audit*

Dopo aver valutato e passato in rassegna i risultati ottenuti dalle distribuzioni dei quattro valori di criterio, analizzando questa fase secondo diverse variabili. I risultati ottenuti mostrano un certo pregiudizio nei confronti dei progetti coordinati da donne. Adesso valuteremo il secondo ed ultimo *step* di perizia sui progetti per controllare un possibile tasso di discriminazione.

In questo caso riprenderemo le tabelle 19 e 20 (“v. sotto-paragrafo 3.3.1”) e vi eseguiremo due operazioni: *slicing* e *roll-up*. La prima operazione riduce la

dimensionalità per SESSO e AMMESSO. La seconda aumenta lo stato di aggregazione della tabella per la variabile FINANZIATO. Ne verrà fatta una per genere.

I risultati per L1 sono i seguenti:

| ContaNumeri - SESSO | AMMESSO | FINANZIATO | Totale Risultato |
|-------------------------|--------------|--------------|------------------|
| | yes | | |
| SESSO | no | yes | |
| F | 45,2% | 54,8% | 100,0% |
| Totale Risultato | 45,2% | 54,8% | 100,0% |

Tabella 37. L1 donne operazione *slicing e roll-up*

| ContaNumeri - SESSO | AMMESSO | FINANZIATO | Totale Risultato |
|-------------------------|--------------|--------------|------------------|
| | yes | | |
| SESSO | no | yes | |
| M | 41,9% | 58,1% | 100,0% |
| Totale Risultato | 41,9% | 58,1% | 100,0% |

Tabella 38. L1 uomini operazione *slicing e roll-up*

Le PT riportano quanto era già stato scoperto dalle analisi precedenti, ovvero che i progetti diretti da donne soffrono di discriminazione.

Invece i risultati per L2 sono i seguenti:

| ContaNumeri - SESSO | AMMESSO | FINANZIATO | Totale Risultato |
|-------------------------|--------------|--------------|------------------|
| | yes | | |
| SESSO | no | yes | |
| F | 61,3% | 38,7% | 100,0% |
| Totale Risultato | 61,3% | 38,7% | 100,0% |

Tabella 39. L2 donne operazione *slicing e roll-up*

| ContaNumeri - SESSO | AMMESSO | FINANZIATO | Totale Risultato |
|-------------------------|--------------|--------------|------------------|
| | yes | | |
| SESSO | no | yes | |
| M | 49,0% | 51,0% | 100,0% |
| Totale Risultato | 49,0% | 51,0% | 100,0% |

Tabella 40. L2 uomini operazione *slicing e roll-up*

Anche in questo caso le PT evidenziano una potenziale discriminazione nei confronti dei progetti amministrati da donne.

Quindi concludendo questo paragrafo notiamo una forte discriminazione nei confronti dei capi progetto donne in particolare nella linea di ricerca L2.

Successivamente aumentando l'aggregazione dei dati tramite la *variabile TIPO_ENTE*, si otterrà un cubo bidimensionale SESSO x AMMESSO x TIPO_ENTE e SESSO come dato di conteggio in percentuale.

| ContaNumeri - SESSO | | AMMESSO | FINANZIATO | Totale Risultato |
|-------------------------|-----------|--------------|--------------|------------------|
| SESSO | TIPO_ENTE | no | yes | |
| F | Altro | | 6,5% | 6,5% |
| | CNR | 3,2% | 3,2% | 6,5% |
| | Univ | 41,9% | 45,2% | 87,1% |
| Totale Risultato | | 45,2% | 54,8% | 100,0% |

Tabella 41(a). L1 con variabile TIPO_ENTE

| ContaNumeri - SESSO | | AMMESSO | FINANZIATO | Totale Risultato |
|-------------------------|-----------|--------------|--------------|------------------|
| SESSO | TIPO_ENTE | no | yes | |
| M | Altro | 2,3% | 11,6% | 14,0% |
| | CNR | 7,0% | 4,7% | 11,6% |
| | Univ | 32,6% | 41,9% | 74,4% |
| Totale Risultato | | 41,9% | 58,1% | 100,0% |

Tabella 41(b). L1 con *variabile TIPO_ENTE*

- L'ente con la maggior percentuale di finanziati risulta essere l'Università seguita da Altro e CNR per entrambe le PT.

Invece i risultati per L2 sono:

| ContaNumeri - SESSO | | AMMESSO | FINANZIATO | |
|-------------------------|-----------|--------------|--------------|-------------------------|
| | | yes | | Totale Risultato |
| SESSO | TIPO_ENTE | no | yes | |
| F | Altro | 6,5% | | 6,5% |
| | CNR | 6,5% | | 6,5% |
| | Univ | 48,4% | 38,7% | 87,1% |
| Totale Risultato | | 61,3% | 38,7% | 100,0% |

Tabella 42(a). L2 donne con *variabile TIPO_ENTE*

| ContaNumeri - SESSO | | AMMESSO | FINANZIATO | |
|-------------------------|-----------|--------------|--------------|-------------------------|
| | | yes | | Totale Risultato |
| SESSO | TIPO_ENTE | no | yes | |
| M | Altro | 1,0% | 5,0% | 6,0% |
| | CNR | 5,0% | 8,0% | 13,0% |
| | Univ | 43,0% | 38,0% | 81,0% |
| Totale Risultato | | 49,0% | 51,0% | 100,0% |

Tabella 42(b). L2 uomini con *variabile TIPO_ENTE*

- L'ente con maggior finanziati risulta essere l'Università seguita da Altro e CNR.
- Nella tab. 42(a) nsi nota che non ci sono finanziati per enti come Altro o CNR.

Nella linea di ricerca L2 l'organo con maggior ammessi è sempre l'Università, dove si nota un grande divario fra progetti diretti da uomini e donne. Quest'ultima caratteristica la si scorge anche nella linea di ricerca dei giovani, ma non risulta così evidente come in questo caso.

Continuando l'ispezione di ammessi e non più nello specifico, si aggiunge ancora un grado di aggregazione grazie alla *variabile SETT_ID_1_LV_1* che delinea le aree di ricerca e studio. Si ottiene così un cubo bidimensionale SESSO x (AMMESSO x TIPO_ENTE x SETT_ID_1_LV_1) e SESSO come dato di conteggio in percentuale.

| ContaNumeri - SESSO | | | AMMESSO | FINANZIATO | Totale Risultato |
|-------------------------|-----------|--------------|--------------|--------------|------------------|
| SESSO | TIPO_ENTE | SETT_ID_1_LV | no | yes | |
| F | Altro | LS | | 6,5% | 6,5% |
| | CNR | LS | | 3,2% | 3,2% |
| | | SH | | | 3,2% |
| | Univ | LS | 19,4% | 19,4% | 38,7% |
| | | PE | 6,5% | 16,1% | 22,6% |
| | | SH | 16,1% | 9,7% | 25,8% |
| Totale Risultato | | | 45,2% | 54,8% | 100,0% |

Tabella 43(a). L1 donne con *variabile SETT_ID_1_LV_1*

| ContaNumeri - SESSO | | | AMMESSO | FINANZIATO | Totale Risultato | |
|-------------------------|-----------|--------------|--------------|--------------|------------------|-------------|
| SESSO | TIPO_ENTE | SETT_ID_1_LV | no | yes | | |
| M | Altro | PE | | 7,0% | 7,0% | |
| | | SH | | 2,3% | 4,7% | 7,0% |
| | CNR | LS | | 4,7% | 4,7% | |
| | | PE | | | 4,7% | 4,7% |
| | | SH | | 2,3% | | 2,3% |
| | Univ | LS | 4,7% | 16,3% | 20,9% | |
| | | PE | 9,3% | 20,9% | 30,2% | |
| | | SH | 18,6% | 4,7% | 23,3% | |
| Totale Risultato | | | 41,9% | 58,1% | 100,0% | |

Tabella 43(b). L1 uomini con *variabile SETT_ID_1_LV_1*

- L'ente Altro mostra che per il settore LS 6,5 % dei progetti diretti da donne è l'unico settore finanziato. Per i progetti diretti da uomini in PE e SH sono i due settore finanziati rispettivamente per il 7 % e il 4,7%.
- Il CNR delinea per SH il 3,2 % finanziato per i progetti coordinati da donne. Mentre nei progetti diretti da uomini viene finanziato PE con il 4,7 % .
- Nell'Università per i progetti diretti da donne si evidenzia che 19,4 % di finanziati per LS, il 16,1 % per PE e il 9,7 % per SH. Mentre per i progetti diretti da uomini si ha il 16,3 % per LS, il 20,9 % per PE e il 4,7 % per SH.

I progetti dei settori PE (*physical sciences and engineering*) e SH (*social sciences and Humanities*) registrano un minor successo oppure pari allo 0 % per i progetti diretti da donne. Questa caratteristica è meno evidente nel settore universitario, dove l'affermazione dei progetti coordinati da donne risulta migliore.

- I progetti diretti da donne ottengono maggiori rifiuti nel campo PE e SH, mentre si esaltano in LS.
- Gli uomini raccolgono meno idoneità nel campo si studio LS, differentemente nei settori PH e SH risultano avere una percentuale di idonei più alta.

| ContaNumeri - SESSO | | | AMMESSO | FINANZIATO | Totale Risultato |
|-------------------------|-----------|--------------|--------------|--------------|------------------|
| SESSO | TIPO_ENTE | SETT_ID_1_LV | no | yes | |
| F | Altro | PE | 6,5% | | 6,5% |
| | | CNR | 3,2% | | 3,2% |
| | Univ | PE | 3,2% | | 3,2% |
| | | LS | 16,1% | 32,3% | 48,4% |
| | | PE | 9,7% | | 9,7% |
| | | SH | 22,6% | 6,5% | 29,0% |
| Totale Risultato | | | 61,3% | 38,7% | 100,0% |

Tabella 44(a). L2 donne con *variabile SETT_ID_1_LV_1*

| ContaNumeri - SESSO | | | AMMESSO | FINANZIATO | Totale Risultato |
|-------------------------|-----------|--------------|--------------|--------------|------------------|
| SESSO | TIPO_ENTE | SETT_ID_1_LV | no | yes | |
| M | Altro | PE | 1,0% | 5,0% | 6,0% |
| | | CNR | 1,0% | 1,0% | 2,0% |
| | Univ | PE | 3,0% | 7,0% | 10,0% |
| | | SH | 1,0% | | 1,0% |
| | | LS | 9,0% | 12,0% | 21,0% |
| | | PE | 18,0% | 21,0% | 39,0% |
| SH | 16,0% | 5,0% | 21,0% | | |
| Totale Risultato | | | 49,0% | 51,0% | 100,0% |

Tabella 44(b). L2 uomini con *variabile SETT_ID_1_LV_1*

- L'ente Altro mostra che per i progetti diretti da donne non c'è nessun progetto finanziato. Mentre per i progetti diretti da uomini si ha il 5% di finanziati nel settore PE.
- Il CNR illustra che per i progetti diretti da donne non c'è nessun progetto finanziato. Invece per i progetti diretti da uomini si ha l'1 % per LS e il 7 % PE di progetti finanziati.
- L'Università ottiene per i progetti coordinati da donne il 32,3 % per LS e il 6,5 % per SH di progetti finanziati. Mentre per i progetti diretti da uomini si ha il 12 % per LS, il 21 % per PE e il 5 % per SH di progetti finanziati.

Confrontando la tab. 43 e la tab. 44 si evince che:

- nei progetti coordinati da donne si nota una percentuale minore o pari allo 0 % per il campo di ricerca SH in tutti e tre gli organi, rispetto ai colleghi maschi.
- nel campo di studio PE si nota una grande differenza in questo caso fra progetti diretti da uomini e donne per idonei e non, per entrambe le linee di ricerca.
- per quanto riguarda il campo LS si nota meno differenza rispetto ai precedenti.

Quindi:

- in entrambe le linee di ricerca la maggior parte dei capi progetto fa riferimento all'Università.
- si evince già un certa divario fra i due sessi in base ai progetti finanziati e non.
- si capisce che i progetti condotti da donne in particolare nei settori SH e PE ottengono molti meno finanziamenti rispetto a quelli dei colleghi maschi.

Nel prossimo punto si valuterà in base all'analisi dei costi e della produttività, le differenze fra progetti coordinati da uomini e donne.

3.3.5 Analisi dei costi e produttività *step 2*

Nel sotto-paragrafo precedente si è valutato e passato in rassegna i risultati ottenuti per l'*audit*, analizzando questa fase secondo diverse variabili. In questo caso riprenderemo le tabelle 25 e 26 (“v. sotto-paragrafo 3.3.2”) per i costi e le tabelle 27 e 28 (“v. sotto-paragrafo 3.3.2”) per la produttività e vi eseguiremo due operazioni: *slicing* e *roll-up*. La prima operazione riduce la dimensionalità AMMESSO. La seconda aumenta lo stato di aggregazione della tabella per la variabile FINANZIATO. I risultati ottenuti mostrano una certa disparità fra i progetti coordinati da uomini e donne. Questo paragrafo si divide in due fasi:

- a) Analisi sui costi dei progetti finanziati e non al fine di valutare una possibile disparità.
- b) Analisi della produttività in base al numero di pubblicazioni e alla media degli autori.

Nella **FASE A** si crea un PT con dimensione SESSO x AMMESSO e con dati la media dei costi totali.

| Media - COSTO_TOT | AMMESSO | FINANZIATO | Totale Risultato |
|-------------------------|----------------|----------------|------------------|
| | yes | | |
| SESSO | no | yes | |
| F | 700.256 | 943.105 | 833.431 |
| M | 1.039.071 | 1.026.962 | 1.032.031 |
| Totale Risultato | 890.839 | 993.020 | 948.834 |

Tabella 45. L1 Fase A

| Media - COSTO_TOT | AMMESSO | FINANZIATO | Totale Risultato |
|-------------------------|------------------|------------------|------------------|
| | yes | | |
| SESSO | no | yes | |
| F | 1.298.440 | 1.102.334 | 1.222.528 |
| M | 965.185 | 1.107.280 | 1.037.653 |
| Totale Risultato | 1.058.300 | 1.106.337 | 1.081.402 |

Tabella 46. L2 Fase A

Nella tab. 45 si nota che la media di M_YES (progetti finanziati con capi progetto uomini) è nettamente superiore a quella di F_YES (progetti finanziati con capi progetto donne) e precisamente la differenza fra le due medie è 83.857. Mentre la media di M_NO (progetti non finanziati con capi progetto uomini) è molto più vicina alla media di F_NO (progetti non finanziati con capi progetto donne) e la differenza fra le due medie è 338.814. Ergo che i progetti ammessi per M_YES hanno un costo medio più alto rispetto a quello di F_YES.

Nella tab. 46 la situazione è ribaltata rispetto alla tab. 45; la media di M_YES è leggermente superiore a quella di F_YES e la differenza fra le due medie è di 4.946. Mentre la media di M_NO è inferiore alla media di F_NO e la differenza fra le due medie è 333.255. Ergo che i progetti ammessi per M_YES hanno un costo medio più alto rispetto a quello di F_YES.

Confrontando L1 e L2 si nota che le due situazioni sono completamente opposte: in L1 si nota un apparente favoritismo di M_YES, ma in L2 si nota un apparente favoritismo per F_YES.

Nella **FASE B** si crea un PT con dimensione SESSO x NUM_PUBB e con dati la media NUM_PUBB.

| Media - NUM_PUBB | AMMESSO | | FINANZIATO | | Totale Risultato |
|-------------------------|---------|-------------|------------|-------------|------------------|
| | no | yes | no | yes | |
| SESSO | | | | | |
| F | | 12,6 | | 11,1 | 11,7 |
| M | | 11,4 | | 15,3 | 13,7 |
| Totale Risultato | | 11,9 | | 13,6 | 12,9 |

Tabella 47. L1 Fase B

| Media - NUM_PUBB | AMMESSO | | FINANZIATO | | Totale Risultato |
|-------------------------|---------|-------------|------------|-------------|------------------|
| | no | yes | no | yes | |
| SESSO | | | | | |
| F | | 15,7 | | 22,3 | 18,3 |
| M | | 20,4 | | 22,0 | 21,2 |
| Totale Risultato | | 19,1 | | 22,1 | 20,5 |

Tabella 48. L2 Fase B

Si nota una maggiore produttività in L2 rispetto ad L1 .

Nel prossimo punto si valuterà in base all'analisi di regressione, da cosa dipendono queste differenze fra progetti coordinati da uomini e donne.

3.3.6 Analisi di regressione

Nel sotto-paragrafo precedente si è valutato e passato in rassegna i risultati ottenuti per lo *step2*, nonché *audit*, esaminando questa fase secondo diverse variabili. I risultati ottenuti mostrano una certa disparità fra capi progetto uomini e donne, in particolare nella seconda linea di ricerca. In quest'ultima fase di analisi si cercherà di capire da cosa dipendono queste differenze studiando le associazioni tra le variabili grazie all'analisi di regressione. Quest'ultima permette di esplorare le relazioni tra due insieme di valori.

Un primo approccio all'analisi di regressione è la creazione di uno *scatter-plot* o

grafico di dispersione. Quest'ultimo fornisce una descrizione visiva della relazione esistente tra le due variabili. Nello specifico i dati sono riportati su un diagramma bidimensionale ed indicati da un punto, le cui coordinate corrispondono ai valori X sull'asse delle ascisse e ai valori Y sull'asse delle ordinate. Lo *scatter-plot* è stato ideato con l'idea di verificare, appunto graficamente, se le coppie di punti (e quindi i fenomeni statistici), presentano una qualche forma di regolarità ed in particolare, per vedere come i punti si disperdono attorno ad un particolare valore. Nel secondo passo si può quantificare il tipo di associazione del grafico di dispersione tramite le misure di associazione. La misura di associazione che utilizzeremo è la correlazione. In statistica per correlazione si intende una relazione tra due variabili casuali tale che a ciascun valore della prima variabile corrisponda con una certa regolarità un valore della seconda. Quindi la correlazione individua una relazione fra due variabili tramite un indice. Quest'ultimo ha un valore pari a 1 se i dati sono esattamente allineati lungo una retta crescente, oppure -1 in caso contrario (la retta è decrescente). A parte il valore 0 che indica assenza di correlazione, negli altri casi i valori tra 0 e 0,3 esprimono una correlazione debole, i valori tra 0,3 e 0,8 esprimono una correlazione moderata, mentre i valori maggiori di 0,8 esprimono una forte correlazione. Quindi come regola empirica si può assumere che valori da 0,8 ad 1 (e rispettivamente da -1 a -0,8) determinano una relazione lineare di tipo diretto (o rispettivamente inverso). Quindi se esiste un'associazione fra due variabili, esse saranno influenzate reciprocamente. Il confronto fra due variabili si effettua solo se fra esse esiste un legame logico, perché la meccanica applicazione delle diverse tecniche può portare a risultati assurdi.

In questo caso si studia che tipo di relazione sussista fra le variabili del *database* FIRB per comprendere le differenze fra i progetti finanziati da donne e uomini. Inizialmente si prenderà in esame la produttività dei progetti che viene rappresentata dal numero di pubblicazioni. Quest'ultima esprime quante pubblicazioni sono state prodotte per un determinato progetto. Nelle tab. 27, 28, 47 e 48 si nota che per entrambe le linee di ricerca i progetti diretti da uomini producono mediamente un numero maggiore di pubblicazioni rispetto ai progetti diretti da donne. Potrebbe darsi che le donne ottengono meno finanziamenti proprio perché producono meno pubblicazioni rispetto agli uomini. Quindi in primo luogo si studierà se il numero di

pubblicazioni influisce sulle valutazioni progetto, tramite la correlazione per comprendere se le differenze fra uomini e donne dipendono da questo fattore. In secondo luogo si analizzerà il costo totale dei progetti. Quest'ultimo esprime un costo totale dato da: numero delle donne facente parte del progetto, numero di giovani ricercatori, ricercatori famosi, più altri costi di gestione. Quindi si studierà se il costo totale dei progetti influisce sulle valutazioni progetto, grazie alla correlazione per comprendere se le differenze fra uomini e donne dipendono da questo fattore.

Ogni associazione sarà studiata in base alla linea di ricerca e in base al sesso del coordinatore progetto, poi sarà calcolata la correlazione e generato il relativo *scatter-plot*.

Si studia che tipo di relazione esiste fra il numero di pubblicazioni e i quattro criteri di valutazione, già citati precedentemente (“v. sotto-par. 2.2.1”) Questo per capire se il numero di pubblicazioni influisce sulle valutazioni del progetto.

I risultati per L1 coordinatori uomini sono i seguenti:

1. NUM_PUBB vs VAL_CRITERIO_1: **0,1** (“v. tab.49(a)”)
2. NUM_PUBB vs VAL_CRITERIO_2: **0,1** (“v. tab.49(b)”)
3. NUM_PUBB vs VAL_CRITERIO_3: **0,1** (“v. tab.49(c)”)
4. NUM_PUBB vs VAL_CRITERIO_4: **0,2** (“v. tab.49(d)”)

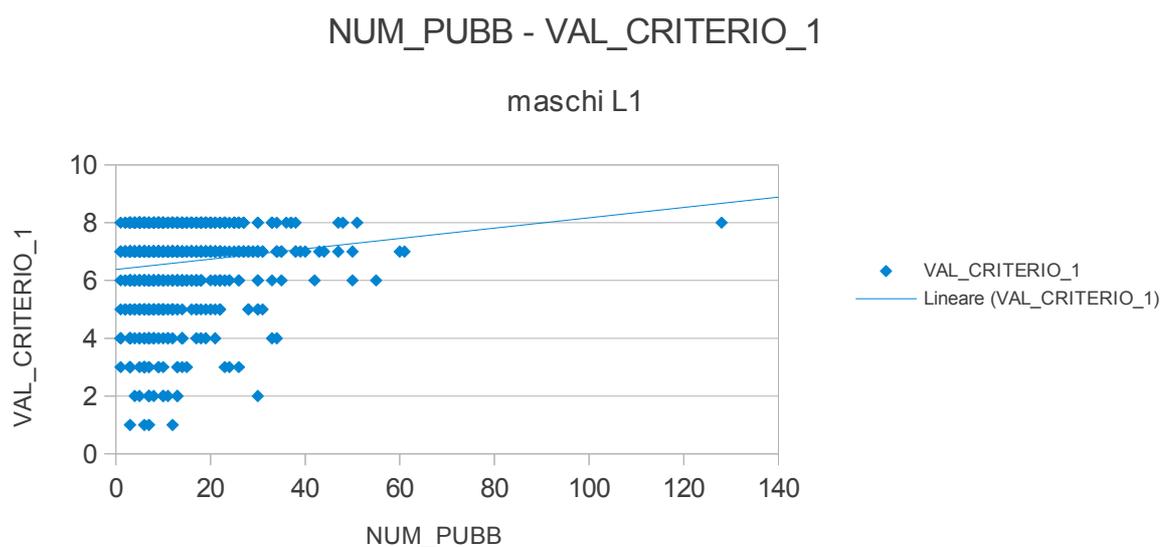


Tabella 49(a). L1 Grafico di dispersione coordinatori uomini, numero pubblicazioni e scientific relevance (score 0-8).

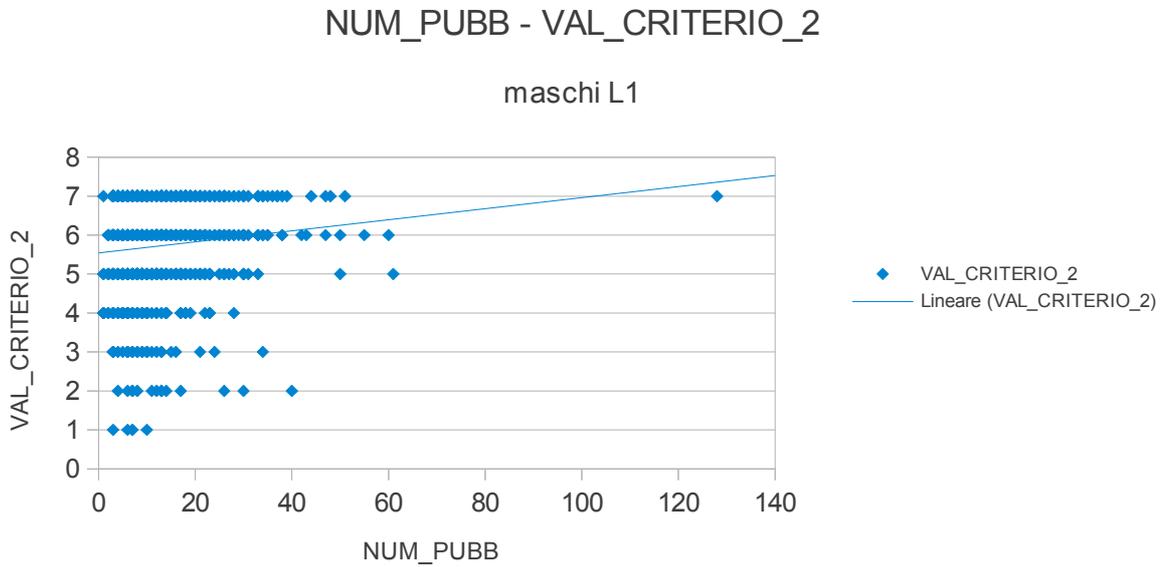


Tabella 49(b). L1 Grafico di dispersione coordinatori uomini, numero pubblicazioni e impact of the project (score 0-7).

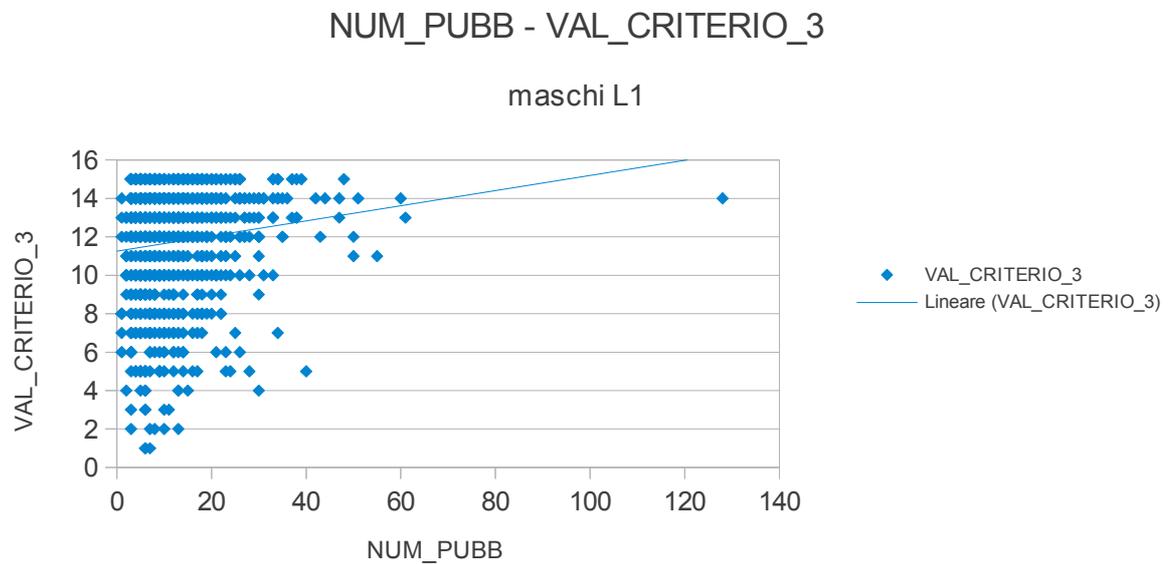


Tabella 49(c). L1 Grafico di dispersione coordinatori uomini, numero pubblicazioni e scientific and technical value (score 0-15)

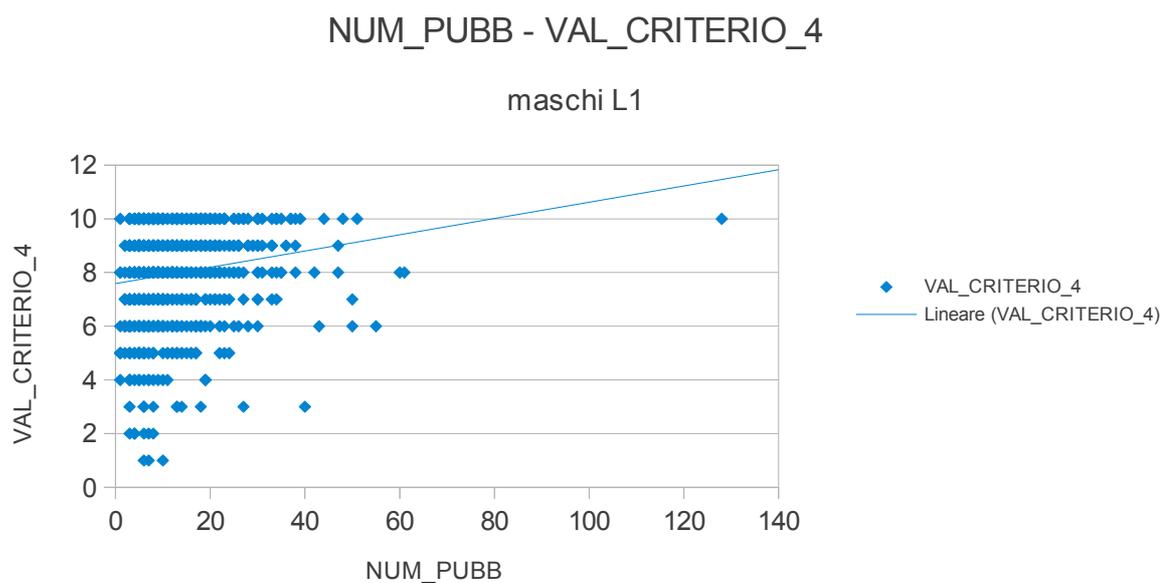


Tabella 49(d). L1 *Grafico di dispersione coordinatori uomini, numero pubblicazioni e quality of the partnership (score 0-10).*

In tutti e quattro i casi l'indice di correlazione tra il numero delle pubblicazioni e i quattro criteri di valutazione è compreso tra 0,1 e 0,2; quindi essendo quasi prossimo allo zero, si può dire che il numero di pubblicazioni non influisce sulle valutazioni ottenute dei progetti, per quanto riguarda il sesso maschile.

I risultati per L1 coordinatrici femmine, sono i seguenti:

1. NUM_PUBB vs VAL_CRITERIO_1: **0,1** (“v. tab. 50(a)”)
2. NUM_PUBB vs VAL_CRITERIO_2: **0,1** (“v. tab. 50(b)”)
3. NUM_PUBB vs VAL_CRITERIO_3: **0,1** (“v. tab. 50(c)”)
4. NUM_PUBB vs VAL_CRITERIO_4: **0,2** (“v. tab. 50(d)”)

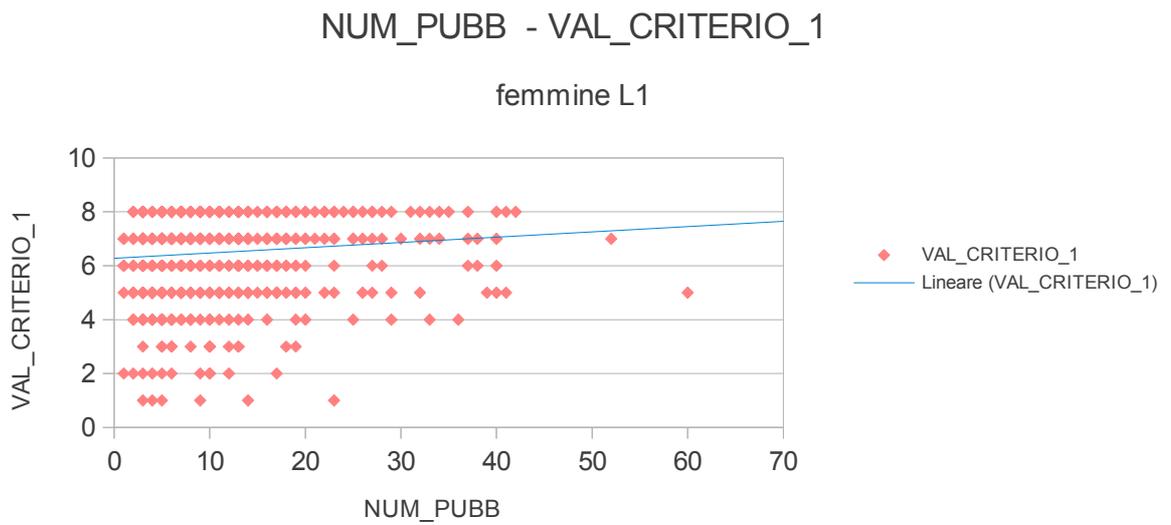


Tabella 50(a). L1 *Grafico di dispersione coordinatrici donne, numero pubblicazioni e scientific relevance (score 0-8).*

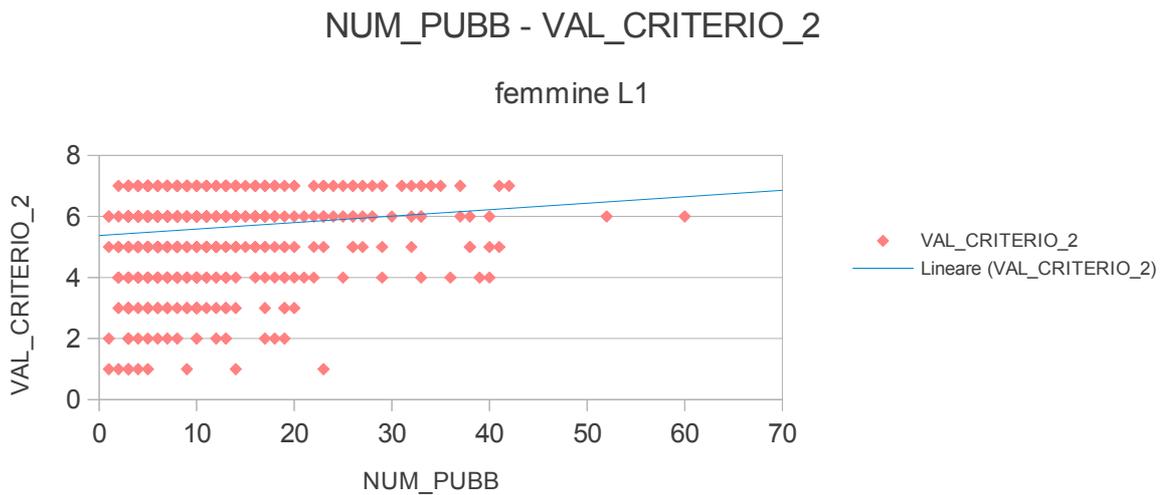


Tabella 50(b). L1 *Grafico di dispersione coordinatrici donne, numero di pubblicazioni e impact of the project (score 0-7).*

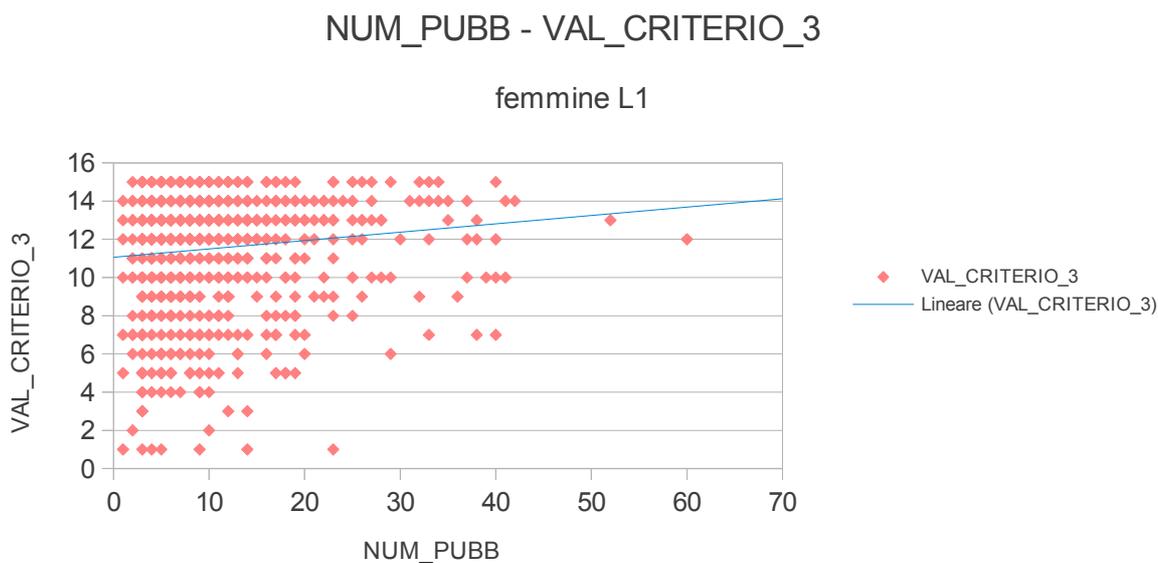


Tabella 50(c). L1 *Grafico di dispersione coordinatrici donne, numero pubblicazioni e scientific and technical value (score 0-15).*

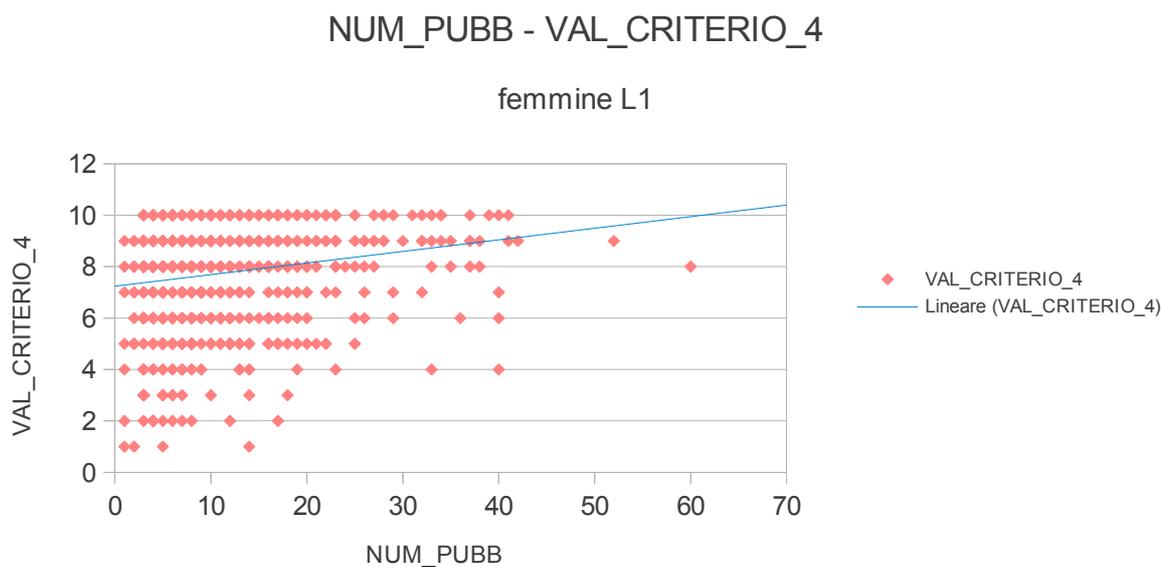


Tabella 50(d). L1 *Grafico di dispersione coordinatrici donne, numero pubblicazioni e quality of the partnership (score 0-10).*

Così come per i coordinatori uomini, anche per le coordinatrici donne sembra non esistere nessun tipo di relazione tra il numero delle pubblicazioni e i quattro criteri di valutazioni, in quanto l'indice di correlazione è prossimo allo zero.

Confrontando i coordinatori uomini e donne di L1 i risultati dell'indice della correlazione mostrano che:

- Per entrambe la linea di ricerca L1 si nota l'assenza di relazione tra le due variabili, sia per il sesso maschile che femminile. Quindi il numero di pubblicazioni non influisce sulle quattro valutazioni dei progetti in L1.

I risultati per L2 coordinatori uomini sono i seguenti:

1. NUM_PUBB vs VAL_CRITERIO_1: **0** (“v. tab.51(a)”)
2. NUM_PUBB vs VAL_CRITERIO_1: **0** (“v. tab.51(b)”)
3. NUM_PUBB vs VAL_CRITERIO_1: **0,1** (“v. tab.51(c)”)
4. NUM_PUBB vs VAL_CRITERIO_1: **0,1** (“v. tab.51(d)”)

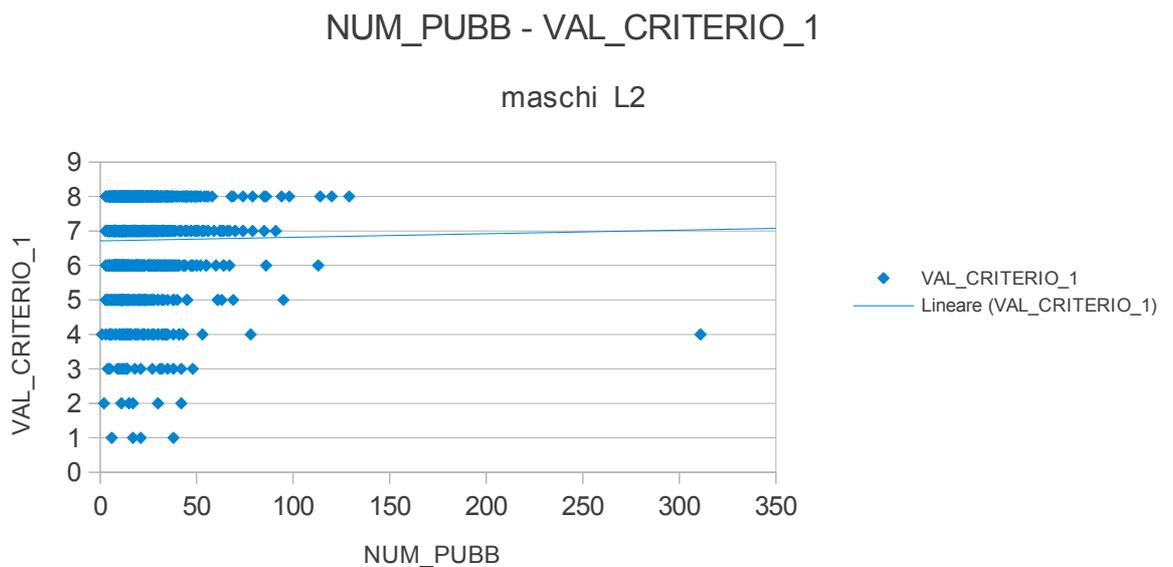


Tabella 51(a). L2 *Grafico di dispersione coordinatori uomini, numero pubblicazioni e scientific relevance (score 0-8).*

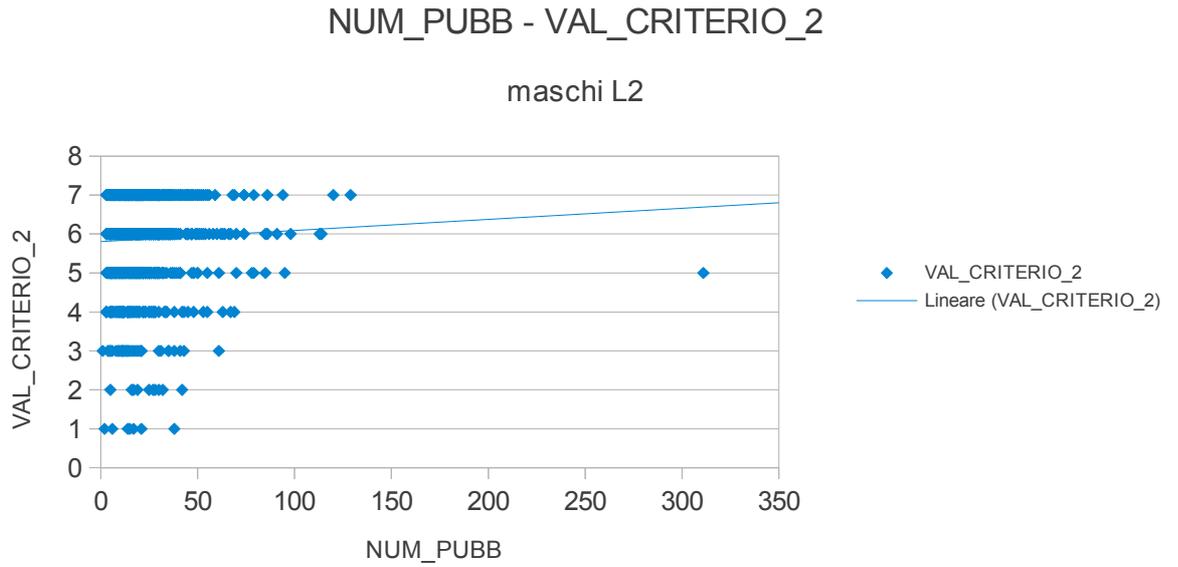


Tabella 51(b). L2 *Grafico di dispersione coordinatori uomini, numero pubblicazioni e impact of the project (score 0-7).*

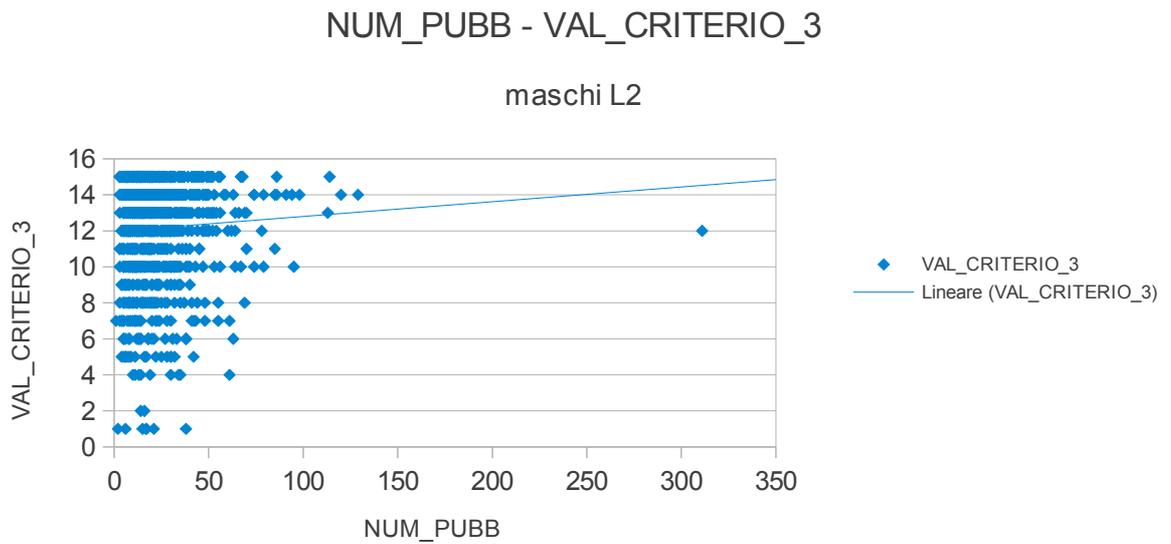


Tabella 51(c). L2 *Grafico di dispersione coordinatori uomini, numero pubblicazioni e scientific and technical value (score 0-15).*

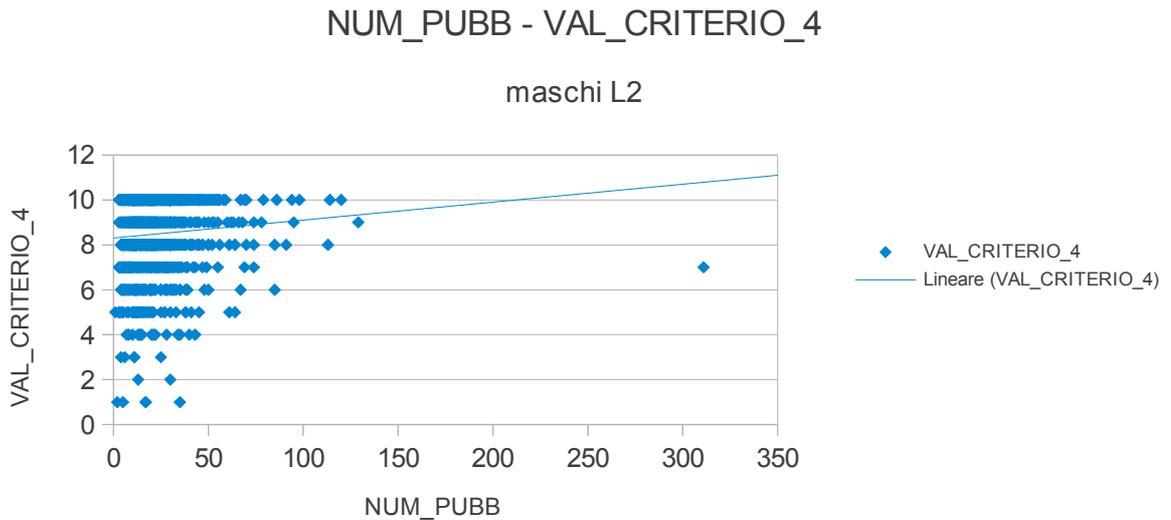


Tabella 51(d). L2 *Grafico di dispersione coordinatori uomini, numero pubblicazioni e quality of the partnership (score 0-10).*

Nella linea di ricerca L2 il numero di pubblicazioni non influisce sulle quattro valutazioni dei progetti, per quanto riguarda il sesso maschile, poiché in tutti e quattro i casi l'indice di correlazione è compreso tra 0 e 0,1.

I risultati per L2 coordinatrici donne sono i seguenti:

5. NUM_PUBB vs VAL_CRITERIO_1: **0,1** (“v. tab.52(a)”)
6. NUM_PUBB vs VAL_CRITERIO_1: **0,1** (“v. tab.52(b)”)
7. NUM_PUBB vs VAL_CRITERIO_1: **0** (“v. tab.52(c)”)
8. NUM_PUBB vs VAL_CRITERIO_1: **0,1** (“v. tab.52(d)”)

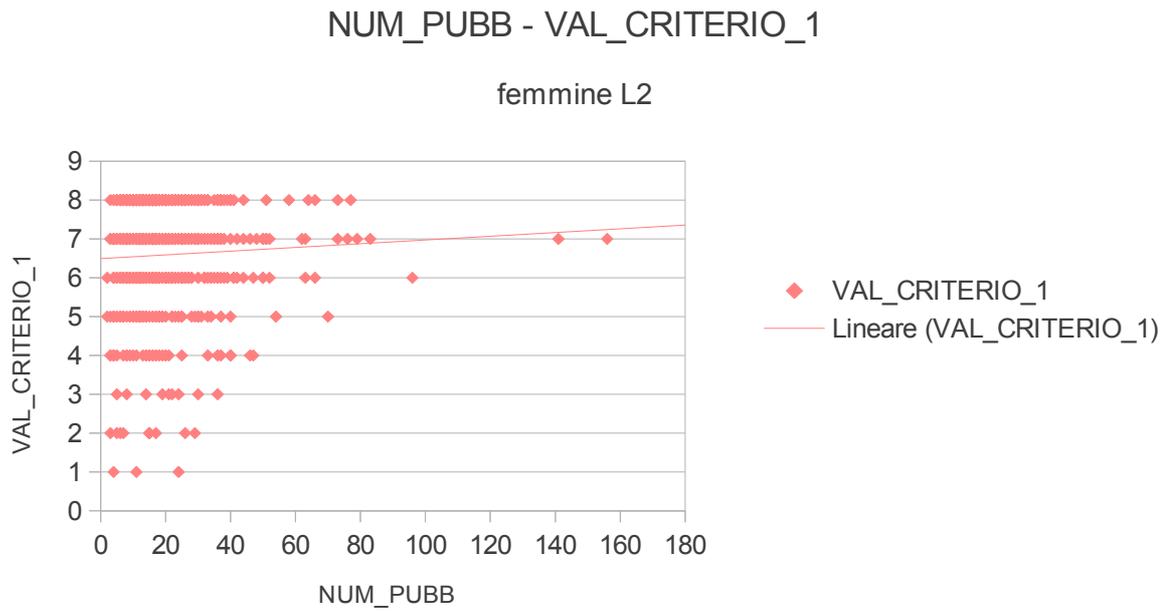


Tabella 52(a). L2 *Grafico di dispersione coordinatrici donne, numero pubblicazioni e scientific relevance (score 0-8).*

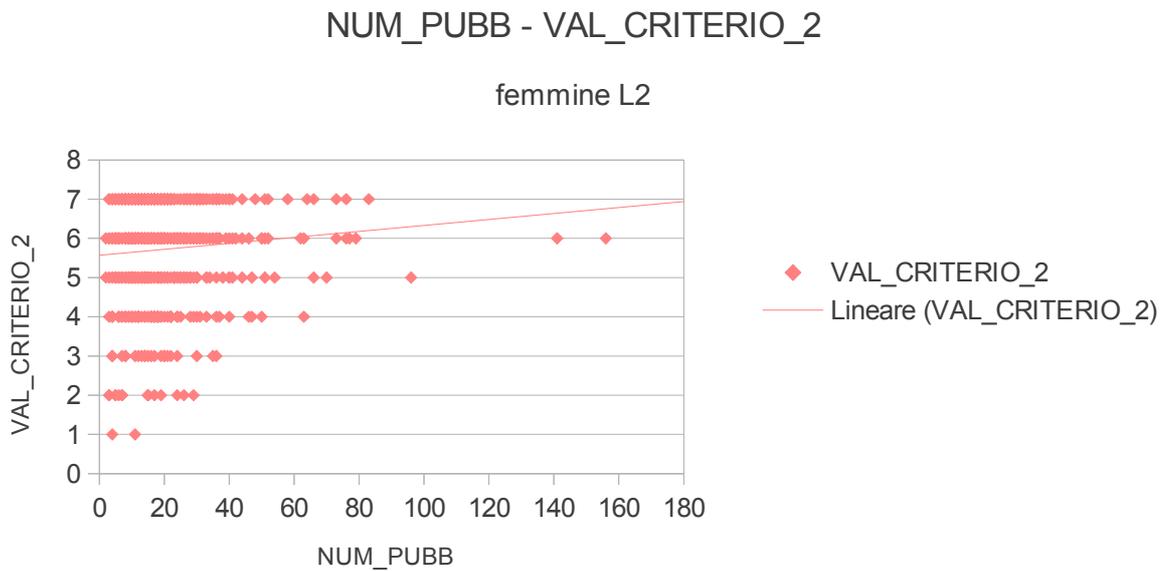


Tabella 52(b). L2 *Grafico di dispersione coordinatrici donne, numero pubblicazioni e impact of the project (score 0-7).*

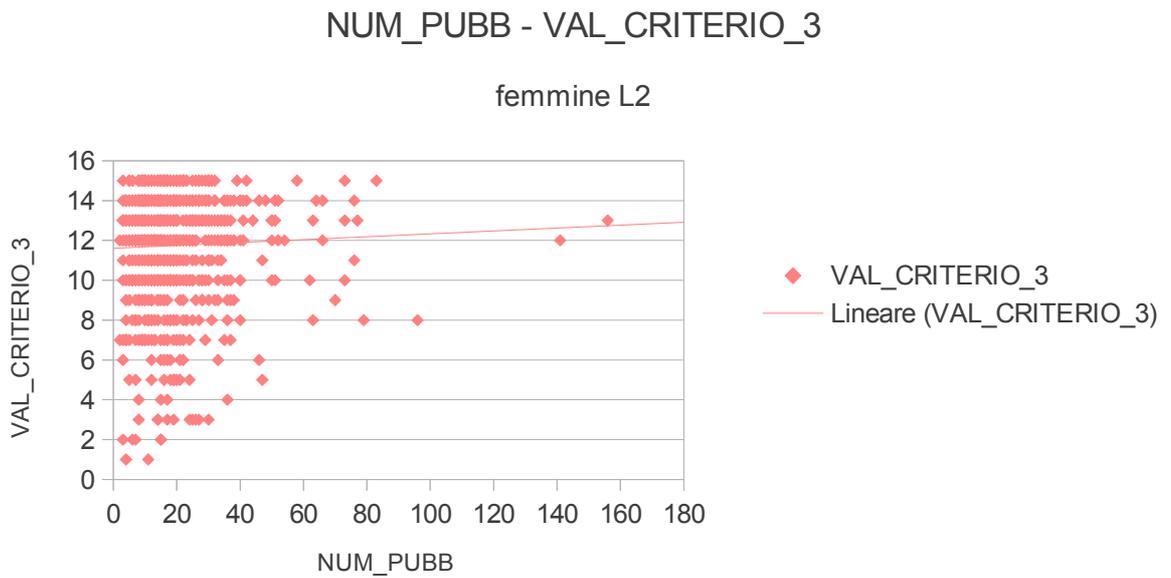


Tabella 52(c). L2 *Grafico di dispersione coordinatrici donne, numero pubblicazioni e scientific and technical value (score 0-15).*

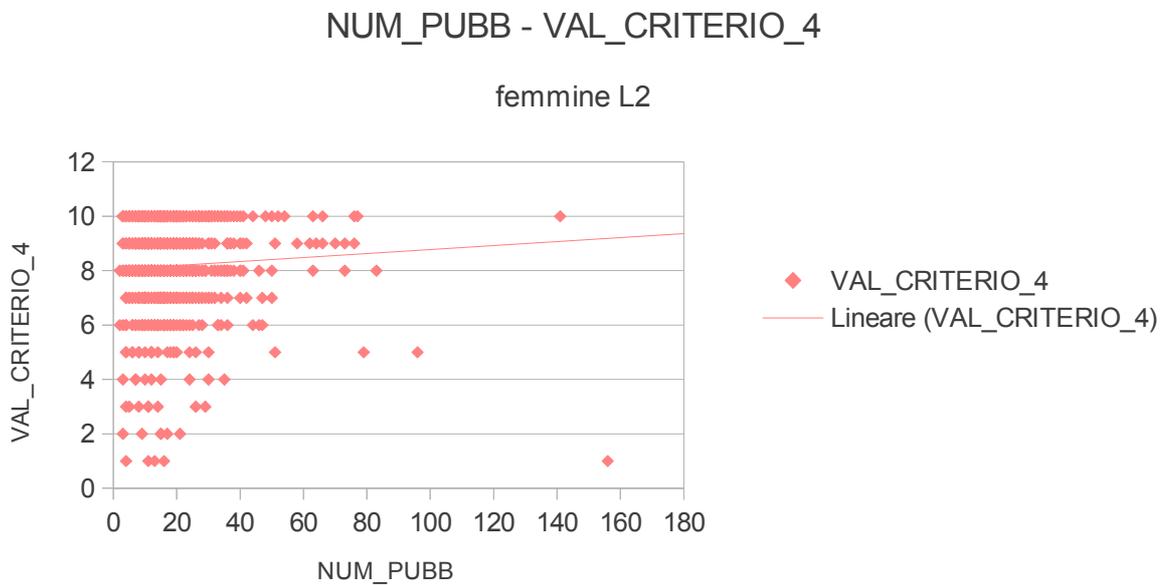


Tabella 52(d). L2 *Grafico di dispersione coordinatrici donne, numero pubblicazioni e quality of the partnership (score 0-10).*

Nella linea di ricerca L2 l'indice di correlazione è prossimo allo zero, quindi il numero di pubblicazioni non influisce sulle valutazioni dei progetti per quanto riguarda il sesso femminile.

Confrontando le due linee di ricerca L1 e L2 , si nota che :

- c'è assenza di correlazione tra il numero di pubblicazioni e i quattro criteri di valutazione sia per progetti diretti da uomini che donne. Questo perché l'indice di correlazione oscilla sempre tra valori compresi tra lo 0 e 0,2 in tutte le correlazioni effettuate.
- in molti casi l'indice di correlazione risulta equivalente sia nei progetti coordinati da uomini che da donne.

Si deduce quindi che il numero di pubblicazioni non influisce sulle valutazioni dei progetti diretti da uomini, né sulle valutazioni dai progetti diretti da donne. Si evince che non ci siano differenze tra i progetti diretti da uomini e donne.

Si studia ora che tipo di relazione esiste fra il costo totale e i quattro criteri di valutazione. Questo per capire se il costo totale dei progetti influisce sulle valutazioni di criterio.

I risultati per L1 coordinatori uomini sono i seguenti:

1. COSTO_TOT vs VAL_CRITERIO_1: **0,1** (“v. tab.53(a)”)
2. COSTO_TOT vs VAL_CRITERIO_2: **0,1** (“v. tab.53(b)”)
3. COSTO_TOT vs VAL_CRITERIO_3: **0,1** (“v. tab.53(c)”)
4. COSTO_TOT vs VAL_CRITERIO_4: **0,2** (“v. tab.53(d)”)

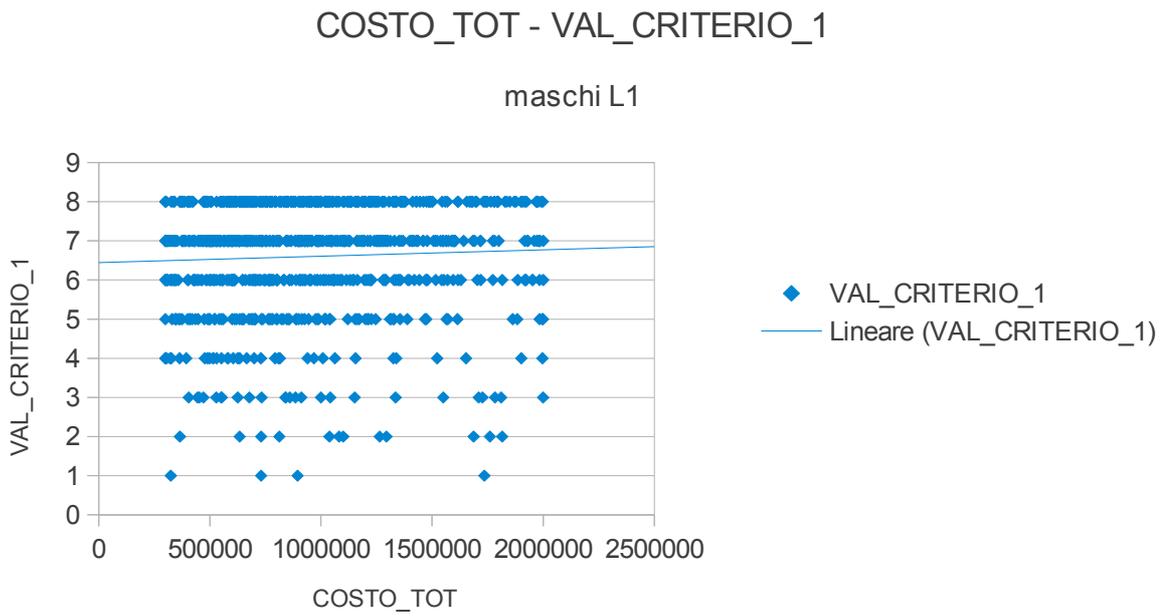


Tabella 53(a). L1 *Grafico di dispersione coordinatori uomini, costo totale e scientific relevance (score 0-8).*

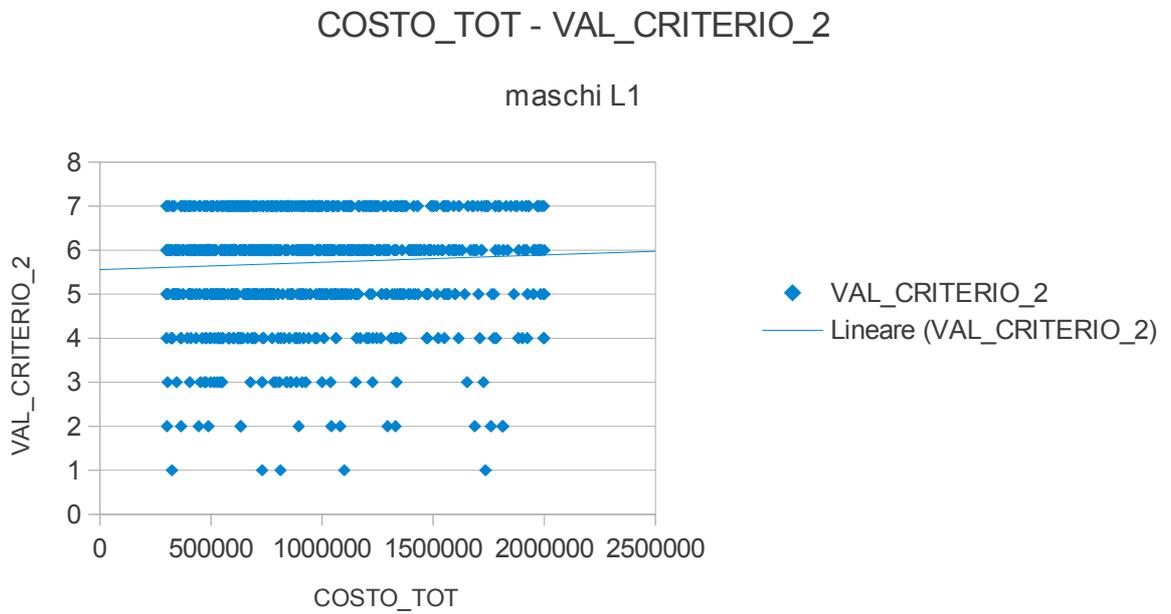


Tabella 53(b). L1 *Grafico di dispersione coordinatori uomini, costo totale e impact of the project (score 0-7).*

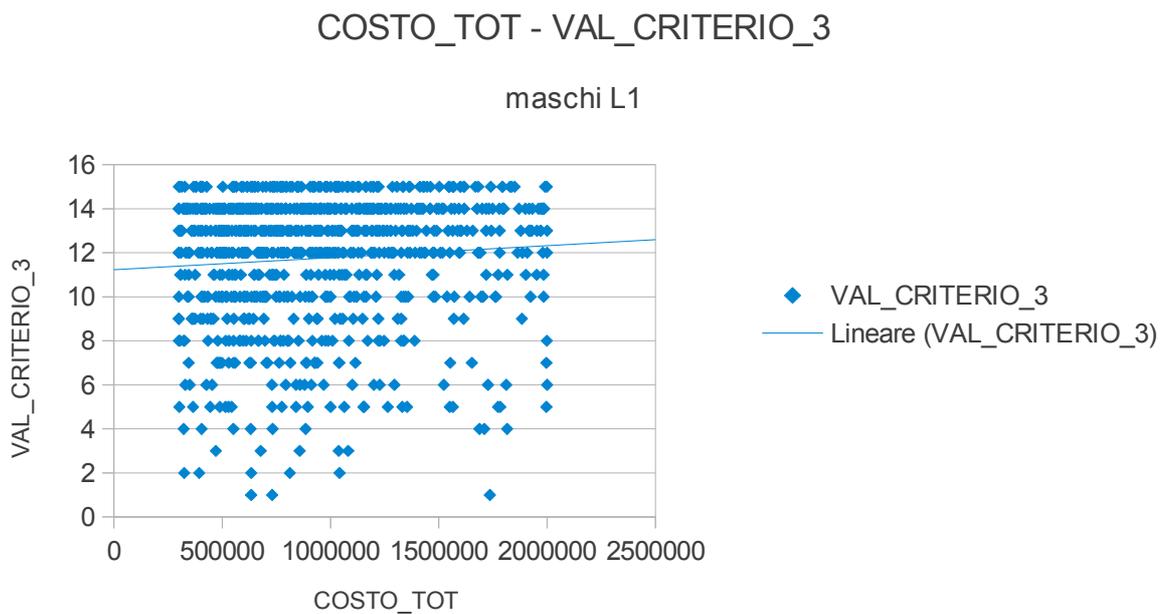


Tabella 53(c). L1 Grafico di dispersione coordinatori uomini, costo totale e scientific and technical value (score 0-15)

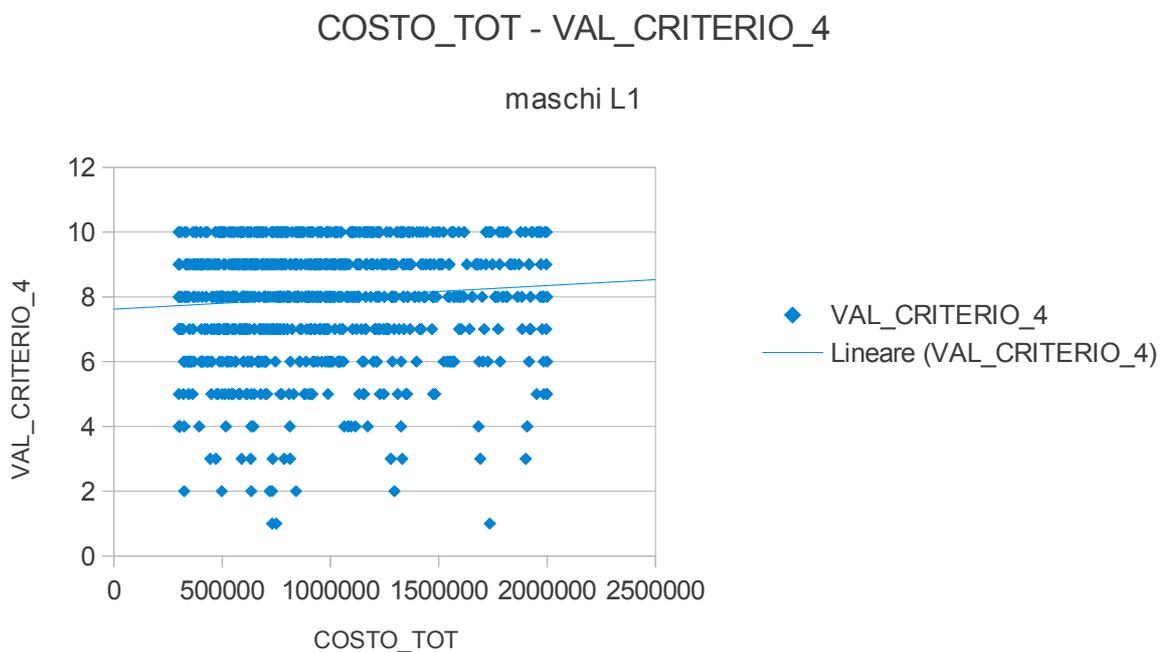


Tabella 53(d). L1 Grafico di dispersione coordinatori uomini, costo totale e quality of the partnership (score 0-10).

I risultati dimostrano che il costo totale non influisce sulle quattro valutazioni dei progetti per i coordinatori uomini nella linea di ricerca L1. Questo perché si hanno indici di correlazione molto deboli che oscillano tra 0,1 e 0,2 .

Risultati per L1 coordinatrici femmine sono i seguenti:

5. COSTO_TOT vs VAL_CRITERIO_1: **0,1** (“v. tab.54(a)”)
6. COSTO_TOT vs VAL_CRITERIO_2: **0,1** (“v. tab.54(b)”)
7. COSTO_TOT vs VAL_CRITERIO_3: **0,1** (“v. tab.54(c)”)
8. COSTO_TOT vs VAL_CRITERIO_4: **0,2** (“v. tab.54(d)”)

I risultati dimostrano che il costo totale non influisce sulle valutazioni dei progetti per i coordinatori uomini di L2. Questo perché si hanno indici che presentano una correlazione debole che oscilla dallo 0,1 allo 0,2.

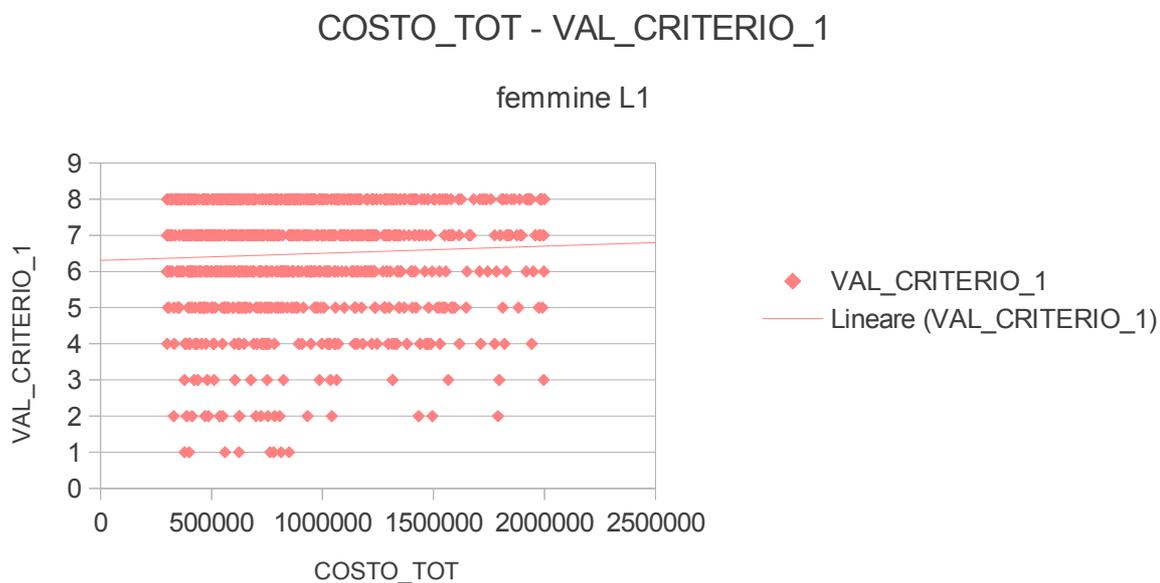


Tabella 54(a). L1 *Grafico di dispersione coordinatrici donne, costo totale e scientific relevance (score 0-8).*

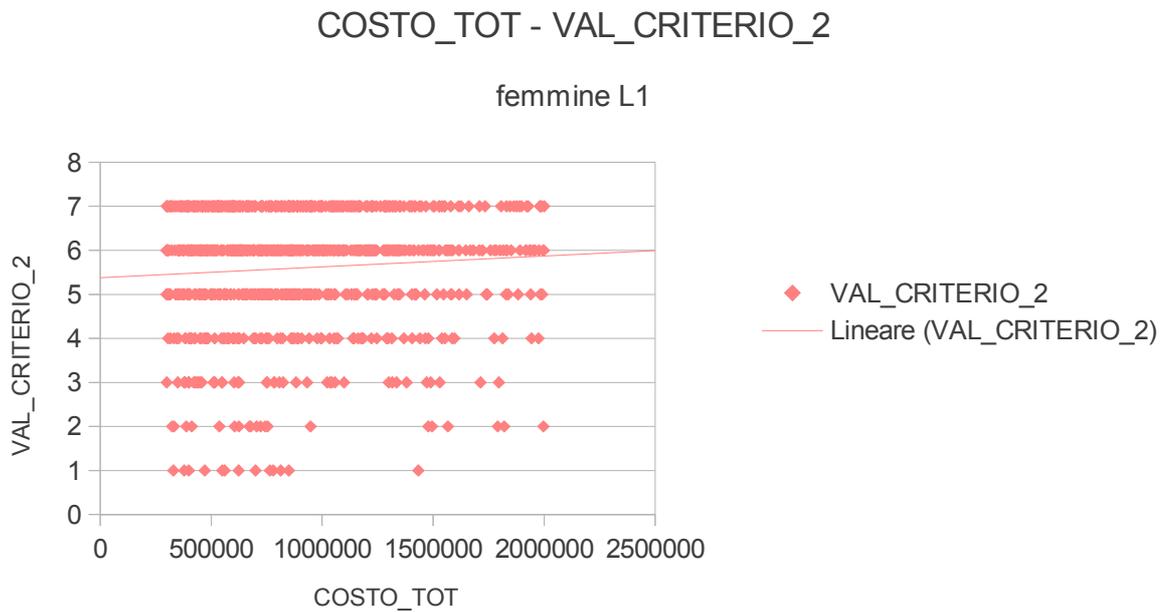


Tabella 54(b). L1 Grafico di dispersione coordinatrici donne, costo totale e impact of the project (score 0-7).

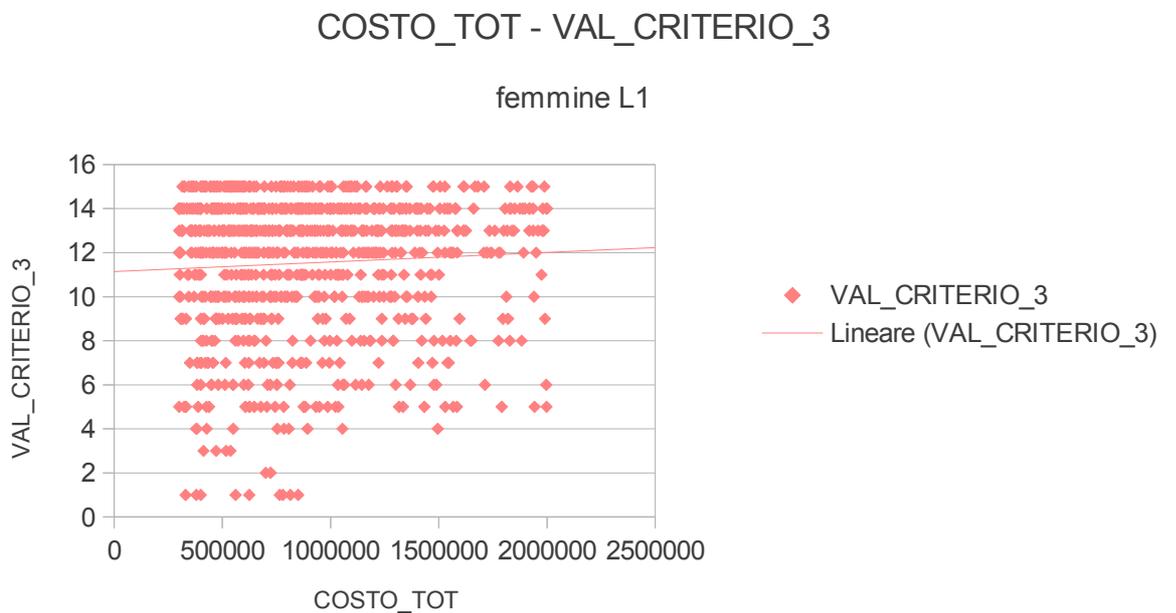


Tabella 54(c). L1 Grafico di dispersione coordinatrici donne, costo totale e scientific and technical value (score 0-15).

COSTO_TOT - VAL_CRITERIO_4

femmine L1

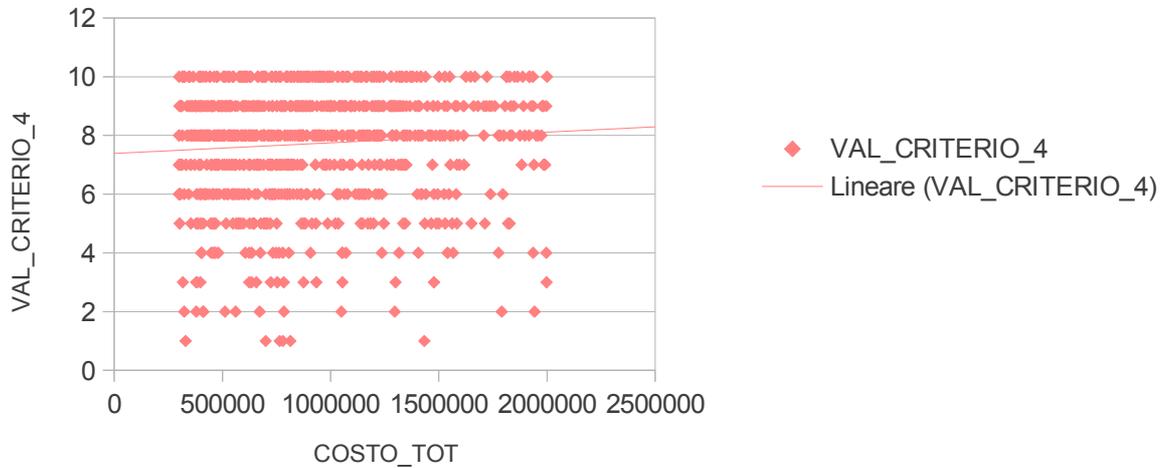


Tabella 54(d). L1 *Grafico di dispersione coordinatrici donne, costo totale e quality of the partnership (score 0-10).*

Nella linea di ricerca L1 per coordinatrici donne si hanno indici di correlazione deboli. Che oscillano tra lo 0,1 e 0,2. Quindi il costo totale non influisce sulle valutazioni dei progetti.

I risultati per i coordinatori uomini in L2 sono i seguenti:

1. COSTO_TOT vs VAL_CRITERIO_1: **0** (“v. tab.55(a)”)
2. COSTO_TOT vs VAL_CRITERIO_2: **0** (“v. tab.55(b)”)
3. COSTO_TOT vs VAL_CRITERIO_3: **0** (“v. tab.55(c)”)
4. COSTO_TOT vs VAL_CRITERIO_4: **0** (“v. tab.55(d)”)

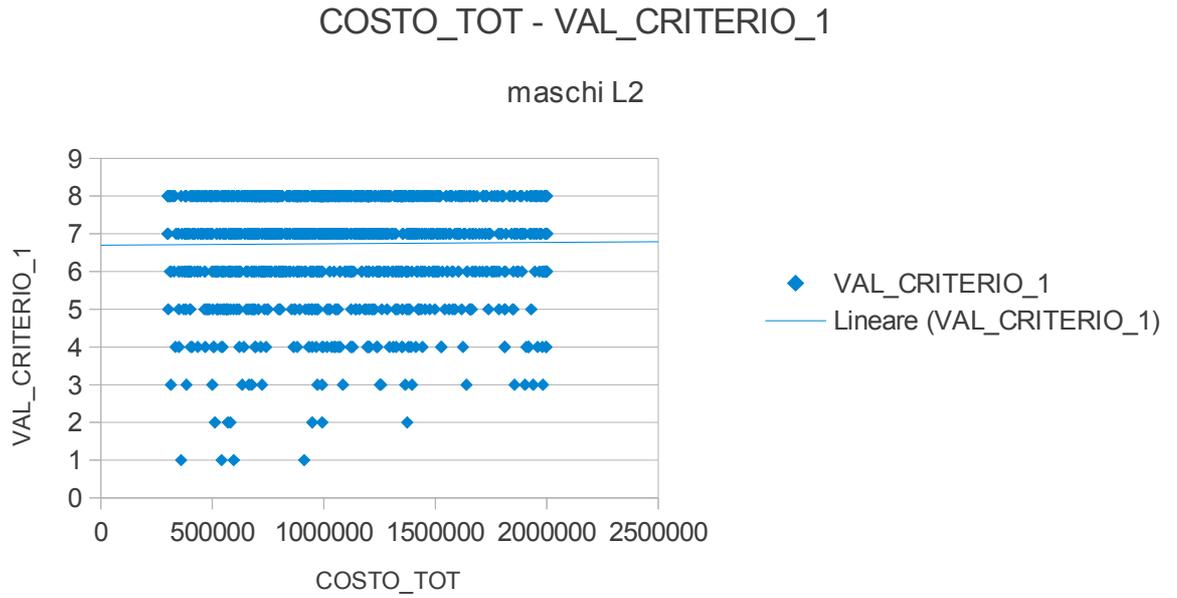


Tabella 55(a). L2 Grafico di dispersione coordinatori uomini, costo totale e scientific relevance (score 0-8).

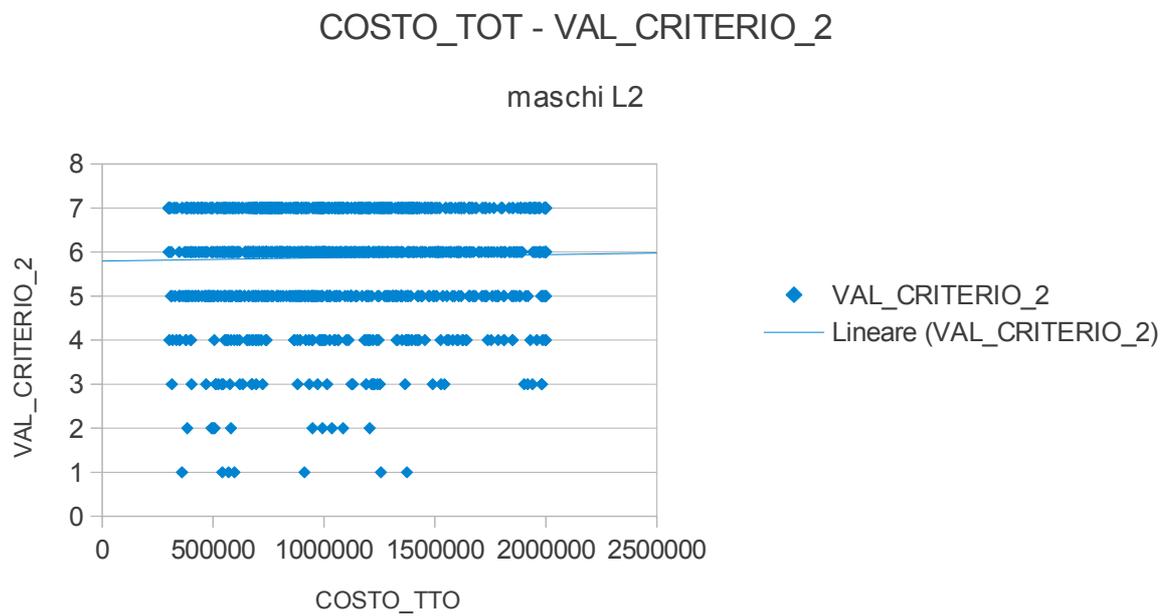


Tabella 55(b). L2 Grafico di dispersione coordinatori uomini, costo totale e impact of the project (score 0-7)

COSTO_TOT - VAL_CRITERIO_3

maschi L2

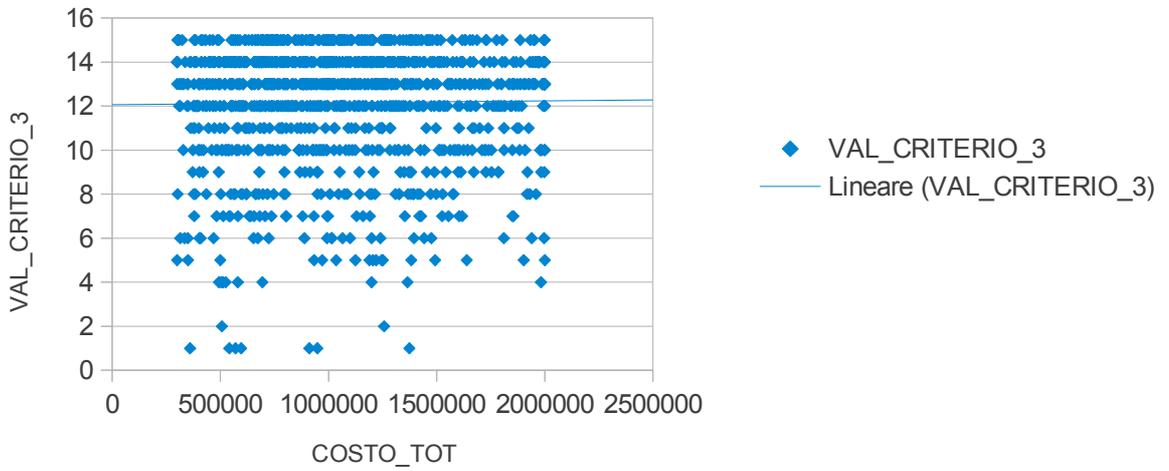


Tabella 55(a). L2 Grafico di dispersione coordinatori uomini, costo totale e scientific and technical value (score 0-15).

COSTO_TOT - VAL_CRITERIO_4

maschi L2

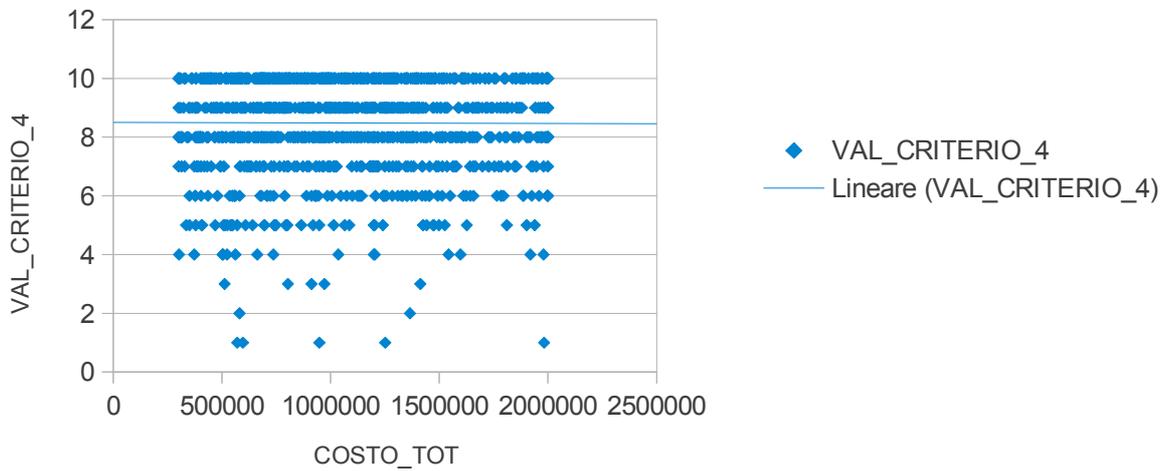


Tabella 55(d). L2 Grafico di dispersione coordinatori uomini, costo totale e quality of the partnership (score 0-10).

I risultati per i coordinatrici donne in L2 sono i seguenti:

5. COSTO_TOT vs VAL_CRITERIO_1: **0** (“v. tab.56(a)”)
6. COSTO_TOT vs VAL_CRITERIO_2: **0** (“v. tab.56(b)”)
7. COSTO_TOT vs VAL_CRITERIO_3: **0** (“v. tab.56(c)”)
8. COSTO_TOT vs VAL_CRITERIO_4: **0** (“v. tab.56(d)”)

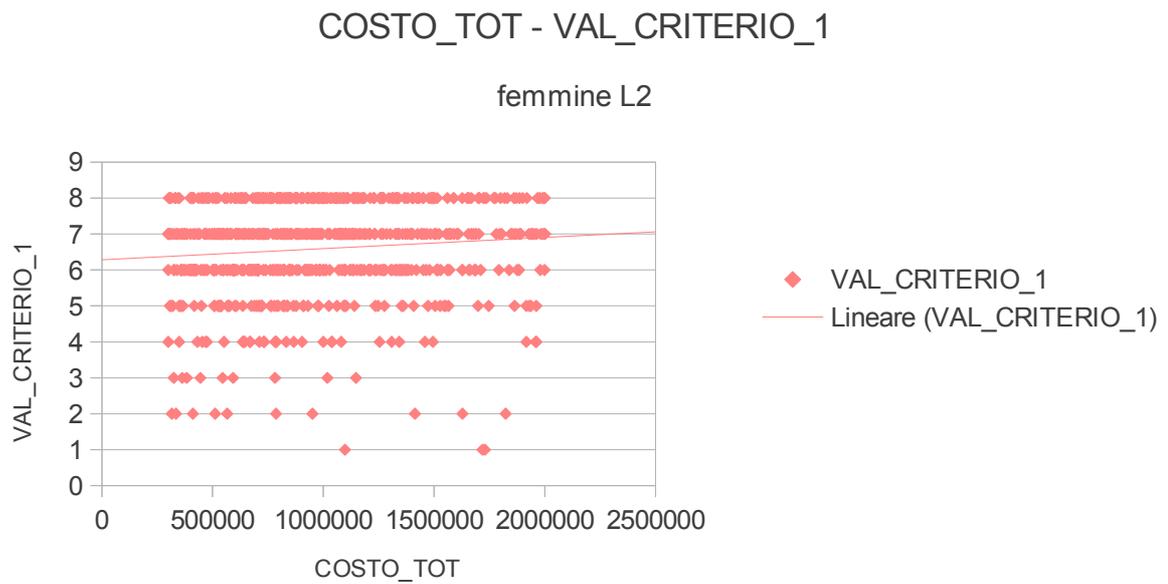


Tabella 56(a). L2 *Grafico di dispersione coordinatrici donne, costo totale e scientific relevance (score 0-8).*

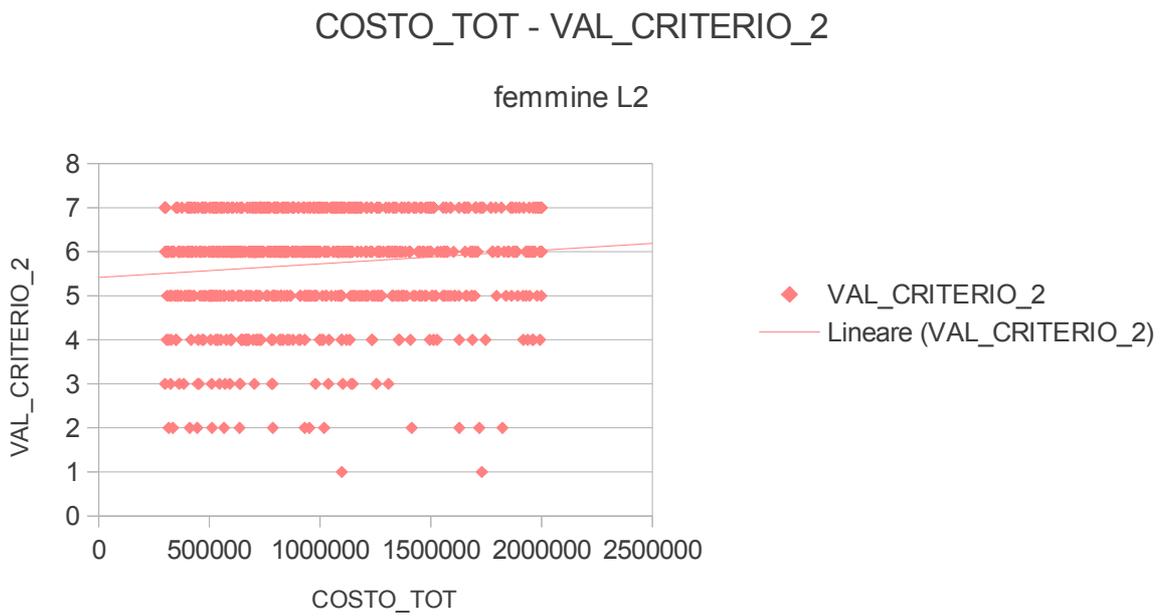


Tabella 56(b). L2 Grafico di dispersione coordinatrici donne, costo totale e impact of the project (score 0-7).

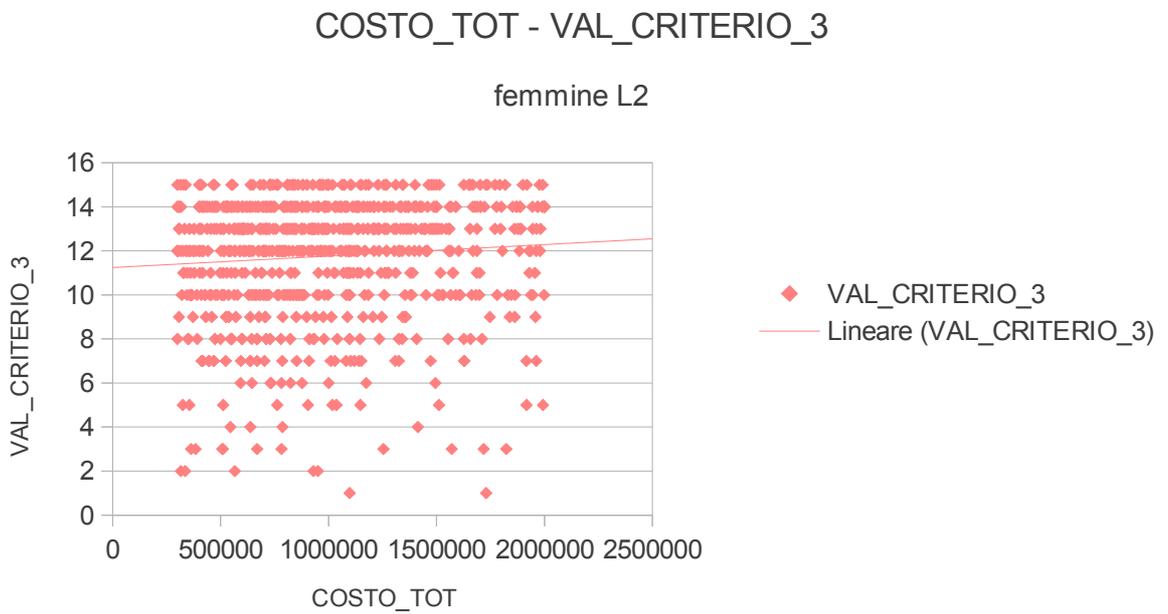


Tabella 55(c). L2 Grafico di dispersione coordinatrici donne, costo totale e scientific and technical value (score 0-15).

COSTO_TOT - VAL_CRITERIO_4

femmine L2

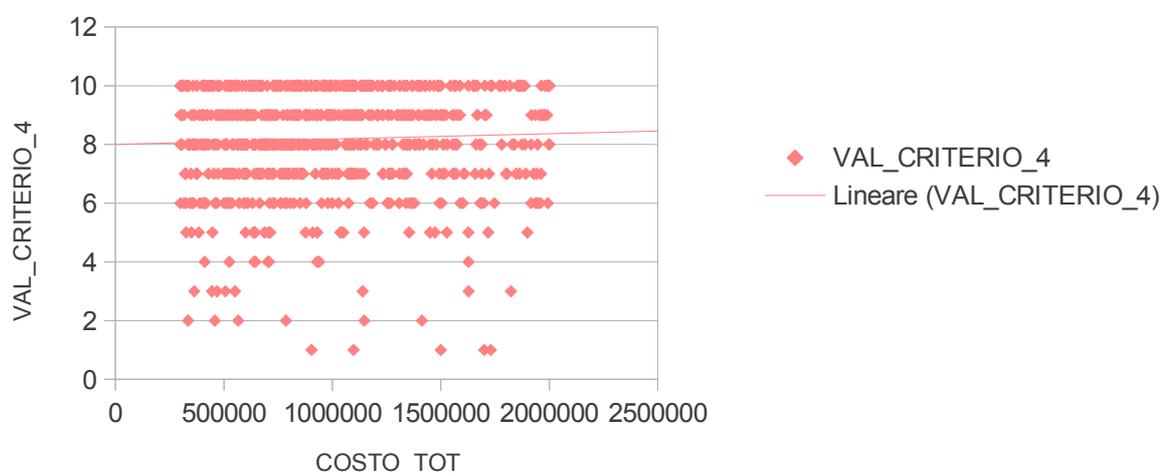


Tabella 55(d). L2 *Grafico di dispersione coordinatrici donne, costo totale e quality of the partnership (score 0-10)*

Dai grafici precedenti nella linea di ricerca L2 l'indice di correlazione tra il costo totale e i quattro criteri di valutazione è uguale a 0, in tutti i casi analizzati. Così si evince che il costo totale dei progetti non influisce sulle valutazioni dei progetti sia per il sesso femminile che per quello maschile.

Confrontando L1 e L2 dai risultati della correlazione si evidenzia che:

- per entrambe le linee di ricerca c'è assenza di relazione tra le due variabili. Quindi il costo totale dei progetti non influisce sulle valutazioni dei progetti.

Concludendo notiamo assenza di correlazione sia per progetti diretti da uomini che donne in entrambe le linee di ricerca (L1-L2). Questo perché si rileva che l'indice di correlazione oscilla in tutti i casi analizzati tra lo 0 e 0,2 effettuate per entrambe le linee di ricerca. Confrontando i risultati ottenuti per entrambe le linee di ricerca in base al sesso, notiamo che in alcuni casi l'indice di correlazione C1 (costo totale e valutazione progetto) e l'indice di correlazione C2 (numero pubblicazioni e valutazione progetto) femminile risulta leggermente più forte di quella maschile, anche se non sono differenze da considerarsi rilevanti.

4. Conclusioni

La *discrimination discovery* è sicuramente un'attività difficile da trattare, se non altro perché, per sua natura, abbraccia diverse discipline: per conoscere il dominio del problema bisogna saper trattare con le varie regolamentazioni date dalla giurisprudenza, oltre al fatto che per dare una buona rappresentazione analitica bisogna avere delle basi di analisi e statistica.

Il lavoro si articola in tre fasi principali: una più teorica con la ricerca e lo studio del materiale in letteratura; la descrizione del set di dati FIRB; la descrizione del modello sviluppato per il trattamento e l'analisi delle variabili nel set di dati FIRB per un lavoro di *discrimination discovery*. Inoltre vuole essere una rassegna dei risultati ottenuti dai processi di analisi di dati sul database FIRB.

Nella fase teorica ci si è innanzitutto concentrati sulla definizione del dominio del problema, tentando di capire cosa significasse realmente il termine discriminazione, con una digressione sia in ambito giuridico che in ambito economico. In particolare ci si è soffermati sullo studio della letteratura economico-statistica riguardante la descrizione di diversi approcci intrapresi da lavori più o meno empirici, volti a costruire modelli di scoperta di discriminazione attraverso teorie statistiche.

Nella seconda fase di descrizione del set di dati FIRB è sembrato opportuno descrivere la struttura e le proprietà relative al *database* FIRB, per comprendere a pieno l'origine, l'importanza e il significato dei dati contenuti, in seguito analizzati nel capitolo 3.

Infine a partire dal paragrafo 3.3 si è esaminato il processo di erogazione fondi per cercare una potenziale discriminazione verso i progetti coordinati da donne. Sono state riportate una serie di *pivot table* (PT) che descrivono le basi statistiche sulle quali si è fondato il progetto di ricerca. Le analisi sono state effettuate in base al sesso dei coordinatori e paragonando in parallelo i risultati ottenuti per L1 e L2. In primo luogo si è analizzato lo *step 1* del processo di erogazione fondi, nonché *international peer-review*. Sono state usate le PT per stimare in proporzione i progetti ammessi e non, e per valutare la produttività dei progetti idonei e non. In seguito sono state effettuate analisi delle distribuzioni sulle quattro valutazioni dei progetti; le medie dei voti differiscono solo per lo 0,1 o lo 0,2 a favore dei progetti diretti da

uomini, non mostrando così grandi disparità da progetti coordinati da donne e uomini, come invece era risultato dalla prima analisi al sotto-paragrafo 3.3.1. I risultati ottenuti da questo primo processo di analisi evidenziano che i progetti coordinati da donne ottengono meno idoneità rispetto a quelli diretti da uomini, in particolare nella seconda linea di ricerca L2. I risultati della produttività dimostrano che i progetti diretti da uomini hanno un numero di pubblicazioni più alto rispetto a quelli diretti da donne; in oltre i costi dei progetti diretti da uomini sono più alti rispetto a quelli diretti da donne nella prima linea di ricerca, al contrario in L2 la situazione è ribaltata.

In secondo luogo si è esaminato lo *step 2* del processo di selezione nonché *audit*. Ci si è avvalsi di PT per stimare in proporzione i progetti finanziati e non e per valutare in proporzione la produttività e i costi dei progetti finanziati e non. I risultati ottenuti da questo secondo processo di analisi hanno evidenziato che i progetti coordinati da donne ottengono meno finanziamenti sui progetti rispetto agli uomini, notando così una potenziale discriminazione nei confronti dei capi progetto donne in particolare nella seconda linea di ricerca. In particolare i progetti coordinati da donne sono svantaggiati nell'erogazione fondi nell'ambito *Physical Science and Engineering* e in percentuale minore nell'ambito *Social Sciences and Humanities*. Si è notato anche che i costi dei progetti diretti da uomini sono mediamente più alti rispetto a quelli delle donne.

A questo punto nel sottoparagrafo 3.3.6 si è voluto comprendere da che cosa derivassero queste differenze rilevate nella *discrimination discovery* tra progetti coordinati da donne e uomini. In particolare ci si è avvalsi dell'analisi di regressione, per cercare di comprendere se queste disparità di risultati, evidenziati nello *step 1* e *step 2*, fossero relazionati ad una certa variabile. Quindi si è studiato le associazioni tra produttività – quattro valutazioni progetti e costo totale – quattro valutazioni progetti, separatamente per entrambe le linee di ricerca e per entrambi i sessi. In particolare ciò ci ha permesso di trarre conclusioni: gli indici di correlazione risultano essere molto bassi, in alcuni casi equivalenti a 0, per cui si può dire che non esiste nessuna associazione tra le variabili esaminate. Così da un lato dalle analisi dei costi e della produttività si evince che i progetti ammessi e finanziati, che sono coordinati da uomini, producono un numero maggiore di pubblicazioni rispetto ai

progetti ammessi e finanziati coordinati da donne. Inoltre i progetti ammessi e finanziati coordinati da uomini hanno costi totali più alti. Quindi si è supposto che ci fosse una dipendenza tra le valutazioni dei progetti e quest'ultime variabili. Dall'altro lato analizzando tramite distribuzioni i voti ottenuti dai progetti si notava che non c'erano grandi disparità come forme di distribuzione e come medie dei voti; in più le distribuzioni delle donne hanno un picco al valore massimo. Però i risultati dell'analisi di regressione non mostrano associazione tra il numero pubblicazioni e i costi totali; quindi si è arrivati a pensare che il sospetto di discriminazione sia generato dalle analisi sulle distribuzioni discusse in precedenza. Così alla fine siamo arrivati a sospettare una potenziale discriminazione nei confronti di progetti coordinati da donne.

Bibliografia

Allbright, S. Christan, Wayne Winston and Christopher Zappe. 2005. *Data Analysis and Decision Making with Microsoft Excel*. Revised Third Edition. USA, South-Western.

Atzeni, Paolo, Stefano Ceri, Stefano Paraboschi, Riccardo Torlone. 2009. *Basi di dati: Modelli e linguaggi di interrogazione*. Terza edizione. Milano, McGraw-Hill.

Golfarelli, Matteo e Stefano Rizzi. 2006. *Introduzione al data warehousing*. In: Matteo Golfarelli e Stefano Rizzi. *Data Warehouse: Teoria e pratica della progettazione*. Milano, McGraw-Hill, cap.1.

Michael Middleton. 2006. *Analisi statistica con Excel*. Milano, Apogeo.

Nelson, Stephen L. 2002. *Working with PivotTables*. In: Stephen L. Nelson. *Excel Data Analysis for Dummies*. Wiley Publishing, cap.4.

Rezzani, Alessandro. 2012. *Business Intelligence: Processi, metodi, utilizzo in azienda*. Milano, Apogeo.

Romei, Andrea and Salvatore Ruggieri. 2012. *A multidisciplinary survey on discrimination data analysis*. Pisa.

Romei, Andrea and Salvatore Ruggieri. 2012. *Discrimination Data Analysis: A Multi-Disciplinary Bibliography*. In: *Discrimination and Privacy in the Information Society*. Vol. 3 of SAPERE. Pisa, cap. 6.

Romei, Andrea, Salvatore Ruggieri and Franco Turini. 2012. *Discrimination Discovery in Scientific Project Evaluation: A Case Study*. Pisa.

Zingarelli, Nicola. 2003. *Lo Zingarelli 2004: Vocabolario della lingua italiana*.
Bologna. Zanichelli Editore.

Siti web consultati:

<http://en.wikipedia.org/wiki/Discrimination>

<http://erc.europa.eu/>

<http://eur-lex.europa.eu/it/treaties/dat/32007X1214/htm/C2007303IT.01000101.htm>

http://europa.eu/legislation_summaries/justice_freedom_security/combating_discrimination/index_it.htm

<http://www.interlex.it/testi/dichuniv.htm>

<http://it.wikipedia.org/wiki/Discriminazione>

<http://it.wikipedia.org/wiki/Pregiudizio>

<http://futuroinricerca.miur.it/>

<http://sociale.regione.emilia-romagna.it/contro-le-discriminazioni/>