

TEORIA E TECNICA DELL'INDAGINE STATISTICA E DEL CAMPIONAMENTO (MATR.DISPARI)

ASSEGNAZIONE ESERCITAZIONE GRAFO

MANUELA SCIONI



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



GRAFO: QUESTIONARIO SUGLI SBOCCHI PROFESSIONALI DEI LAUREATI

1 CARATTERISTICHE SOCIODEMOGRAFICHE

1.1 Sesso: Maschio....1☐

Femmina....2☐

1.2 Et  (in anni compiuti):

1.3 Citt  di residenza: (men  tendina-classificazione con codici ISTAT)

1.4 Citt  in cui vive attualmente: (men  tendina-classificazione con codici ISTAT – estero=888)

1.5 Cittadinanza: (men  tendina-classificazione con codici ISTAT se ha pi  cittadinanze oltre a quella italiana indicare Italiana)

1.6 Stato civile:

1☐ Celibe/nubile

2☐ Coniugato/a, Convivente

3☐ Separato/a o Divorziato/a

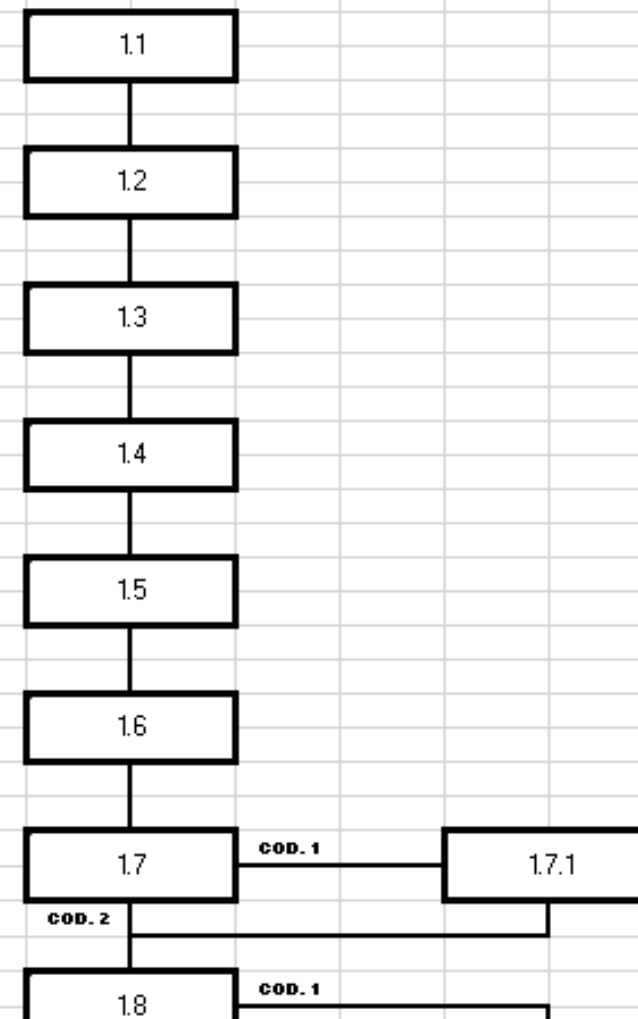
4☐ Vedovo/a

1.7 Ha figli?

1. ☐ S , inserire il numero 1.7.1

2. ☐ No

SEZIONE A. CARATTERISTICHE SOCIODEMOGRAFICHE

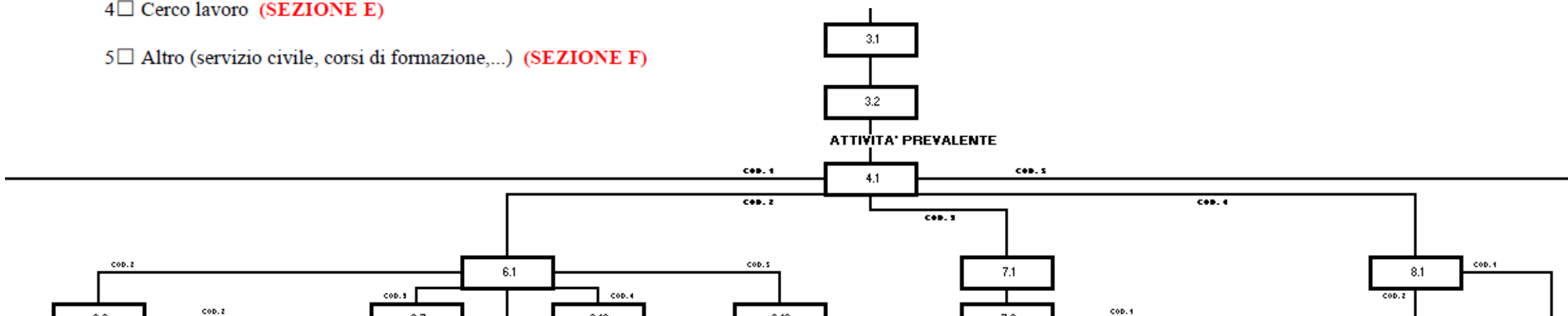


GRAFO: QUESTIONARIO SUGLI SBOCCHI PROFESSIONALI DEI LAUREATI

4 ATTIVITÀ PREVALENTE

4.1 Attualmente, qual'è la sua attività prevalente?

- 1 ☐ Lavoro (SEZIONE B)
- 2 ☐ Ho proseguito gli studi (corsi di durata superiore a 1 anno) (SEZIONE C)
- 3 ☐ Sto svolgendo uno Stage/Tirocinio (SEZIONE D)
- 4 ☐ Cerco lavoro (SEZIONE E)
- 5 ☐ Altro (servizio civile, corsi di formazione,...) (SEZIONE F)



GRAFO: QUESTIONARIO SUGLI SBOCCHI PROFESSIONALI DEI LAUREATI

- Realizzare il grafo del questionario sugli sbocchi professionali dei laureati
- Inviarlo via mail a manuela.scioni@unipd.it entro il 20 Aprile
- La soluzione sarà caricata su moodle dopo tale data
- Se avete dubbi, non esitate a contattarmi!

TEORIA E TECNICA DELL'INDAGINE STATISTICA E DEL CAMPIONAMENTO (MATR.DISPARI)

PIANO DI CHECK

MANUELA SCIONI



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



CONTESTO DI RIFERIMENTO

1. Controllo a **posteriori**: base di dati “grezzi”, registrati successivamente alla fase di rilevazione (es. PAPI – Paper Assisted Personal Interview)
2. Controllo a **priori**: intervista con rilevatore o senza, con registrazione dei dati contemporanea alla rilevazione

Fig. 1 – Esempio di base di dati grezza

Mario	Rossi	M	34	3	4	0	1	0	0
Peppe	Giallo	M	45	4	2	1	0	0	1
Rosa	Rosae	F	20	2	4	0	0	0	1
.....									
.....									
.....									
.....									

SITUAZIONI DA VERIFICARE

- **Valori fuori dominio:** il valore di una variabile non appartiene ad un insieme predefinito di valori ammissibili. I valori fuori dominio danno origine ad **errori**
- **Incompatibilità** fra risposte: i valori di una o più variabili contraddicono predefinite regole di natura logica e/o relazioni di tipo matematico. Le incompatibilità portano a **situazioni di errore** o ad **accertamenti**, ma spesso non si sa su quale/i variabile intervenire
- **Valori anomali:** una unità è **anomala** (outlier) quando presenta caratteristiche significativamente diverse da quelle della maggior parte delle unità. I valori anomali danno origine ad **accertamenti** → **VERIFICA SOLO A POSTERIORI**

TIPOLOGIE DI CONTROLLI

- **Controlli di validità o di range:** verificano che i valori assunti da una data variabile siano interni all'intervallo di definizione della variabile stessa.
- **Controlli di consistenza:** verificano che prefissate combinazioni di valori assunti da variabili rilevate in una stessa unità soddisfino certi requisiti (regole di incompatibilità).
- **Controlli statistici (a posteriori):** utilizzati per isolare quelle unità statistiche che presentano, per alcune delle variabili in esse contenute, valori che si discostano in modo significativo dai valori che le stesse variabili assumono nel resto delle unità. Questi valori sono con alta probabilità errati, ma sono necessarie ulteriori verifiche

PIANO DI CHECK

I sopracitati controlli vengono utilizzati per la costruzione di un «piano di check», definito come:

un insieme di vincoli non ridondanti e non contraddittori che devono essere contemporaneamente soddisfatti da ogni unità statistica affinché l'informazione corrispondente possa essere considerata corretta

Le regole che compongono un piano di incompatibilità possono essere distinti in:

- **regole formali**, che derivano dalla struttura del modello, cioè direttamente dalle norme di compilazione e dai "percorsi interni" del modello;
- **regole sostanziali**, che derivano da considerazioni di tipo statistico - matematico, o da conoscenze specifiche a priori del fenomeno oggetto di rilevazione.

TEORIA E TECNICA DELL'INDAGINE STATISTICA E DEL CAMPIONAMENTO (MATR. DISPARI) CONTROLLI A POSTERIORI

MANUELA SCIONI



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



CONTROLLO A POSTERIORI: IDENTIFICAZIONE DEGLI ERRORI

Si procede alla localizzazione dei valori anomali e dei valori sospetti. Tre possibili approcci:

- **Per variabile:** la presenza per una variabile di parecchi valori anomali dovuti a risposte errate è spesso un efficace indicatore di presenza di errore sistematico per quella variabile.
- **Per unità statistica:** la presenza di un'unità statistica caratterizzata da parecchi errori potrebbe indicare mancata attenzione da parte del rispondente
- **Monitoraggio complessivo:** per individuare la tipologia di errore più frequente

MONITORAGGIO SULLE VARIABILI PIÙ PROBLEMATICHE

- Attraverso un confronto fra archivio dei dati grezzi e puliti, si conta il numero di volte in cui la variabile X assume valori diversi all'interno dei due archivi.
- Le variabili alle quali sono associati i valori più alti sono quelle più problematiche, per le quali bisogna fare delle riflessioni per migliorarne la qualità.

MONITORAGGIO SULLE UNITÀ PIÙ PROBLEMATICHE

- Confronto dello stesso record nei due archivi (dati grezzi vs dati puliti)
- Conteggio del numero di variabili che assumono valori diversi con riferimento allo stesso record nei due archivi
- Le unità con un numero maggiore di variabili diverse sono quelle più problematiche

MONITORAGGIO SUGLI ERRORI PIÙ FREQUENTI

- Valutazione complessiva della qualità dei dati e delle tipologie prevalenti d'errore
- Creazione di variabili (inizializzate a 0), tante quante sono le tipologie di controllo. La variabile Check si incrementa di un'unità ogni volta che il check verifica l'errore
- In questo modo si ottiene il numero di volte in cui ogni specifico errore si è presentato, ovvero le situazioni maggiormente soggette a errore

CORREZIONE DEGLI ERRORI

1. Abbiamo trovato i record contenenti valori errati
2. Sappiamo quali sono le variabili responsabili
3. Che modifiche apportiamo per riportare il record nella condizione di accettabilità rispetto ai criteri (piano di incompatibilità o piano di localizzazione dei valori anomali) utilizzati?
 - a) Ritorno alla fonte
 - b) Imputazione deterministica (assegnare un valore al posto del dato errato o mancante, sulla base di regole logiche) o probabilistica (assegnare un valore estratto da un insieme di unità simili a quella su cui bisogna intervenire)
 - c) Trattamento del dato come *item non response*, si lascia il dato mancante

RITORNO ALLA FONTE

- Se possibile, richiedere il dato direttamente al rispondente o a chi ha compilato il questionario
- È dispendioso in termini di tempo e lavoro, ma garantisce di avere il dato più corretto

IMPUTAZIONE DETERMINISTICA O PROBABILISTICA

- Significa assegnare un valore alla variabile al posto del valore errato, sulla base di altre informazioni.
- Tipicamente le informazioni devono essere disponibili all'interno dell'insieme di dati e della stessa unità
- Generalmente, per procedere con sicurezza all'imputazione, è necessario che le informazioni provengano da più di una variabile, altrimenti non sapremmo decidere, fra due variabili con valori incompatibili, qual è quello giusto

N.B.: non affrontiamo in questo corso tutta la tematica dell'imputazione di dati mancanti, che richiederebbe diverse ore di lezione.

TRATTAMENTO DEL DATO COME ITEM *NON RESPONSE*

Se non ci sono strumenti per recuperare o imputare il dato corretto, rimane solo da cancellare il dato errato e trattare quel campo come un dato mancante

TEORIA E TECNICA DELL'INDAGINE STATISTICA E DEL CAMPIONAMENTO (MATR. DISPARI)

IMPOSTAZIONE A PRIORI

MANUELA SCIONI



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



RISOLVERE IL PROBLEMA ALLA FONTE

- **Cosa fare:** Spostare il controllo dei dati il più possibile vicino alla fase di raccolta delle informazioni presso le unità, in modo da rendere più facile il reperimento di informazioni corrette.
- **Come fare:** Tecnologie per l'integrazione del controllo e correzione dei dati con le fasi di intervista o di registrazione, in modo da eliminare, o in ogni caso minimizzare, gli errori attribuibile ad errori di compilazione o registrazione dei modelli
- **Vantaggio:** Alcune tipologie di errori vengono corrette contemporaneamente alla fase di intervista o di registrazione, producendo una migliore qualità finale dei dati ed un risparmio nei tempi e nei costi connessi alle fasi successive di controllo dei dati.

EDITING CONTEMPORANEO ALLA RILEVAZIONE: ESEMPI

Rilevazione CAPI (Computer Assisted Personal Interview)

Rilevazione CATI (Computer Assisted Telephonic Interview)

Rilevazione CAWI (Computer Assisted Web Interview)

Registri amministrativi informatizzati

CONTROLLI DI RANGE

A. CARATTERISTICHE SOCIODEMOGRAFICHE

Età

! Verificare il dato inserito

65

? Indichi l'età in anni compiuti

Età

! Valore troppo basso, si prega di correggere

20

? Indichi l'età in anni compiuti

CONTROLLI DI INCOMPATIBILITÀ

B. PERCORSO FORMATIVO

Tipo di laurea conseguito

❗ l'età è troppo bassa per il tipo di laurea, si prega correggere

- ☐ Triennale
- ☒ Magistrale
- ☐ Ciclo Unico

VARIABILI OBBLIGATORIE

- Alcune variabili (non tutte!) sono irrinunciabili, e la loro compilazione è obbligatoria. Il codice identificativo (es. codice fiscale) è un tipico campo obbligatorio.

1

*Attualmente lei

❗ Domanda obbligatoria

❗ Se si sceglie 'Altro, specificare:', specificare la scelta fatta nel campo testo di accompagnamento.

☐ Lavora

1

☐ Studia

☐ Lavora e studia

☐ Altro, specificare:

NUMERO MASSIMO DI RISPOSTE

Perché ha deciso di svolgere un'attività lavorativa così poco coerente con gli studi svolti? (fino a **due risposte)*

❗ Scegliere una o più delle seguenti opzioni

❗ **Selezionare al massimo 2 risposte**

- ☐ Per le ottime prospettive di carriera
- ☒ Per gratificazione personale
- ☐ Per la necessità di guadagno
- ☒ Per l'alta remunerazione
- ☒ È stato il primo lavoro che ho trovato, non ci ho pensato su troppo
- ☐ Per il desiderio di orientarmi verso una carriera in un settore diverso da quello del titolo di studio
- ☐ Per consuetudine con la professione già esercitata