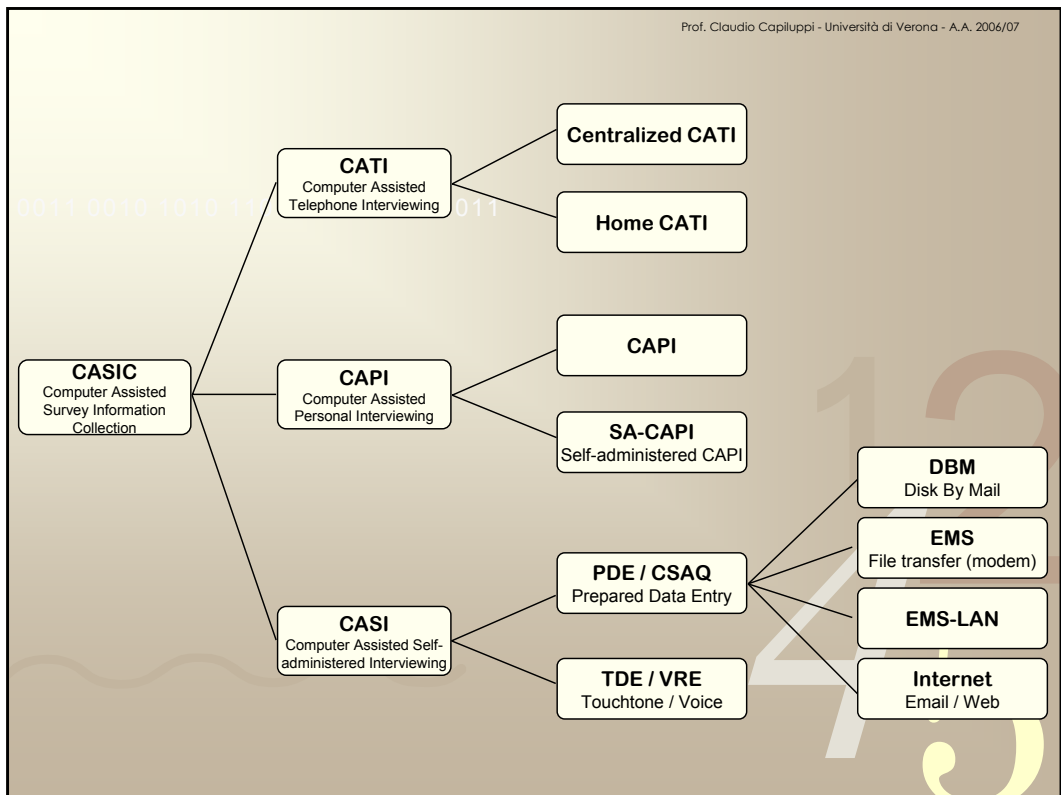


Computer Assisted Interviewing

0011 0010 1010 1101 0001 0100 1011

Prof. Claudio Capiluppi



IL QUESTIONARIO COMPUTER ASSISTED

- L'acronimo *CASIC* (*Computer Assisted Survey Information Collection*) coniato nel 1990 è il termine generalmente adottato per indicare l'intera famiglia delle nuove tecniche di indagine per condurre rilevazioni statistiche assistite mediante computer e software specializzato.
- *CASIC* si aggiunge ai vari termini usati in precedenza, come *CAI* (*Computer Assisted Interviewing*) e *CAQI* (*Computer Assisted Questionnaire Interviewing*), e ricomprende tutte e tre le principali tecniche di indagine *CA*:
 - *CATI*: intervista telefonica con intervistatore che opera con il supporto di un computer
 - *CAPI*: intervista faccia a faccia con intervistatore dotato di un computer portatile/palmare
 - *CASI*: questionario autocompilato dal rispondente direttamente tramite un computer.
- Nell'ultima categoria rientrano tutte le tecniche che non prevedono il ricorso ad intervistatori, qualunque sia il mezzo utilizzato per raggiungere il rispondente:
 - *PDE* (*Prepared Data Entry*): il rispondente usa direttamente sul proprio PC il programma, un questionario informatizzato, per fornire le risposte
 - *TDE* (*Touchtone Data Entry*) e *VRE* (*Voice Recognition Entry*): il rispondente telefona ad un computer e risponde alle domande sintetizzate dal software mediante i tasti del telefono, o mediante riconoscimento vocale

IL QUESTIONARIO COMPUTER ASSISTED

- La tecnica *PDE*, detta anche recentemente *CSAQ* (*Computerized Self Administered Questionnaire*), prevede diverse opzioni operative in funzione della modalità di invio del programma e di raccolta dei file contenenti le risposte:
 - *DBM* (*Disk By Mail*): i rispondenti ricevevano il programma su floppy tramite posta, e sempre per posta rispediscono il floppy con il file delle risposte.
 - *Download/upload file via modem*: il programma viene *downloaded* tramite modem dal computer centrale dell'ente che esegue l'indagine, e ad intervista ultimata il file delle risposte viene *uploaded* ancora via modem.
 - *LAN*: i rispondenti hanno accesso al programma tramite una rete locale.
 - *Email*: i rispondenti ricevono una form in qualche modo compilabile, in allegato ad una email
 - *CAWI* (*Computer Assisted Web Interviewing*): implementazione del *CASI on-line* via Internet, i rispondenti hanno accesso al questionario tramite un Web browser.

IL QUESTIONARIO COMPUTER ASSISTED

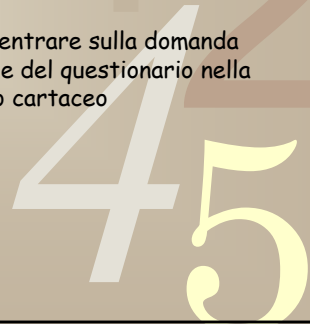
- Un questionario computer-assisted (QCA) è un programma software che:
 - permette di condurre l'intervista dei soggetti, sostituendo al questionario cartaceo una interfaccia interattiva
 - registra le risposte direttamente su supporto magnetico, in una struttura dati opportuna
 - permette di effettuare controlli di correttezza e coerenza sui dati durante l'intervista
 - Un QCA ha dunque una duplice natura:
 - Statistica: è lo strumento di rilevazione dell'indagine, progettato per produrre dati statistici che devono rispondere ad obiettivi conoscitivi prefissati
 - Informatica: è un programma software di una certa complessità, assimilabile ad applicazioni di gestione dati (database)
- ⇒ Richiede al ricercatore una competenza adeguata in entrambe le discipline.

IL QUESTIONARIO COMPUTER ASSISTED

- Un sistema CAI è uno strumento software realizzare indagini CA:
 - permette di ridurre la competenza informatica richiesta al ricercatore per realizzare una indagine CA
 - dovrebbe permettere di affrontare adeguatamente le problematiche di natura statistica in senso lato che possono presentarsi nella progettazione e implementazione del QCA
- Strumenti software per sviluppare questionari computer-assisted:
 - linguaggi di programmazione general-purpose (es. C, Visual Basic)
 - database programmabili (es. Access)
 - ambienti di sviluppo appositamente orientati per realizzare QCA:
 - sistemi language-based, con un proprio linguaggio Q-oriented (es. Blaise, Case/Q, Ci3, QAL, QSL)
 - sistemi con modalità di definizione del questionario non language-based: ambienti visuali o comunque interattivi (es. V-QSL, Captor)

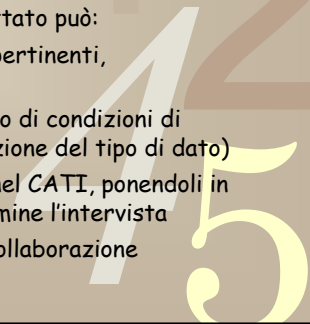
IL QUESTIONARIO COMPUTER ASSISTED

- Differenze principali rispetto al questionario cartaceo:
 - gestione automatica del processo di interrogazione dell'intervistato (sequenza domande, domande filtro)
 - possibilità di realizzare questionari con percorsi più complessi
 - possibilità di costruire domande personalizzate basate sulle risposte precedenti
 - possibilità di rilevare incoerenze durante l'intervista e operare una conciliazione immediata col rispondente
 - diverse caratteristiche dell'interfaccia del questionario (dimensioni video, font caratteri disponibili, visualizzazione delle domande)
 - effetto "segmentazione": l'intervistatore si può concentrare sulla domanda corrente, ma non riesce ad avere una visione d'insieme del questionario nella sua interezza, come quando ha in mano il questionario cartaceo



IL QUESTIONARIO COMPUTER ASSISTED

- Vantaggi dell'indagine CAI: migliore qualità dei dati statistici prodotti
 - riduzione fonti di errore extra-campionario ⇒ accuratezza (precisione)
 - eliminazione errori di percorso
 - correzione degli errori al momento dell'intervista, invece di rimandarli a indagine conclusa, quando l'intervistato non è più disponibile
 - riduzione tempi post-rilevazione ⇒ tempestività
 - integrazione della fase di registrazione dei dati su supporto informatico, per la successiva elaborazione
 - integrazione in fase di rilevazione di controlli di coerenza ⇒ semplificazione della fase di revisione dei dati
- Problemi (banali...): un sistema non adeguatamente progettato può:
 - saltare domande pertinenti o entrare in sezioni non pertinenti, interrompersi a metà strada ...
 - non permettere di registrare un dato (es. per effetto di condizioni di controllo inconsistenti, o anche di errori nella definizione del tipo di dato)
 - ▼ mettere in difficoltà gli intervistatori, soprattutto nel CATI, ponendoli in situazioni da cui è difficile riuscire a condurre a termine l'intervista
 - ▼ mettere a disagio i rispondenti, pregiudicandone la collaborazione



IL QUESTIONARIO COMPUTER ASSISTED

- Un QCA integra aspetti e operazioni normalmente implementate in diverse fasi dell'indagine:
 - la fase di realizzazione del questionario, sostituito dall'interfaccia QCA
 - la fase di organizzazione dei dati su supporto magnetico, cioè la progettazione di una adeguata struttura dati per memorizzare le risposte
 - la fase della revisione dei dati raccolti, cioè la realizzazione del sistema di controlli di coerenza
 - in una certa misura, anche la fase di sintesi ed elaborazione delle informazioni (report sintetici, tabulazione frequenze).
 - lo sviluppo del QCA richiede più tempo e attenzione di quello cartaceo.
- Il ricorso ad un QCA richiede una progettazione accurata e un altrettanto attento pre-test:
 - l'efficacia del QCA è completamente determinata dalla validità della progettazione
 - in fase di progettazione è necessario prevedere tutte le possibili situazioni in cui l'intervistatore (o l'intervistato) possono venire a trovarsi
 - la gestione automatica dell'intervista può diventare il punto critico, perchè non c'è flessibilità, in caso di imprevisti, per "improvvisare" (come nel cartaceo)
 - disegnare il grafo completo del questionario e percorrere in fase di test ogni singolo ramo testando tutti i possibili input

IL QUESTIONARIO COMPUTER ASSISTED

- L'operatività con l'interfaccia QCA risulta più rigida rispetto alla gestione di un questionario cartaceo:
 - la flessibilità dell'interfaccia QCA dipende dalla sua programmazione software quindi, per quanto ben progettata, è limitata a quello che è stato previsto
 - bisognerebbe riuscire a prevedere tutte le possibili situazioni e condizioni che si potranno incontrare durante l'intervista
 - bisogna supportare un modo per acquisire una risposta anche in caso situazioni non previste (es: non so, non risponde, altro->specificare)
 - è opportuno prevedere sempre la possibilità di inserire annotazioni alla domanda corrente.
- Tutte le operazioni/funzioni supportate dal programma QCA devono avvenire in tempo reale durante la conduzione dell'intervista, cioè senza tempi di attesa che possono infastidire l'intervistato (soprattutto nell'indagine CATI):
 - gli spostamenti tra le domande del questionario devono risultare agevoli e veloci: il sistema deve permettere di ritornare con facilità alle domande precedenti, permettere di rivedere ed eventualmente correggere le risposte già fornite
 - il sistema deve gestire in modo consistente l'eventuale modifica di una risposta ad una domanda filtro, che può invalidare una parte ("ramo") del questionario compilato in precedenza
 - la routine di gestione delle incompatibilità rilevate dovrebbe permettere la conciliazione dell'ultima risposta con le altre implicate nella regola

IL QUESTIONARIO COMPUTER ASSISTED

- Il questionario CA è un nuovo e potente strumento nelle mani del ricercatore:
 - consente un maggiore controllo sul processo di intervista
 - richiede anche un maggior investimento iniziale in fase di progettazione
- Nella progettazione del QCA è necessario tenere presenti tutte le indicazioni per la realizzazione di un questionario tradizionale (PAPI): il QCA non riduce la competenza di natura statistica necessaria per la progettazione dello strumento di rilevazione.
- In più, la realizzazione di una buon QCA, anche utilizzando un sistema CAI, richiede che il progettista abbia una sufficiente comprensione:
 - di come si realizza un programma ben strutturato
 - dei principi guida per la progettazione concettuale di un modello entità-relazione.
- È necessario considerare il questionario non solo come una sequenza di domande, ma anche come un *data model*: occorre quindi operare una progettazione concettuale del questionario, come per un database

45

LA PROGETTAZIONE DEL QUESTIONARIO

- La progettazione concettuale del questionario è un modo diverso di considerare il questionario statistico:
 - non solo come una sequenza di domande con una logica di somministrazione,
 - ma anche come un *data model*, cioè un insieme di *entità* e di *relazioni*.
- Il progettista del questionario ha anche il compito di disegnare lo schema concettuale del questionario idoneo per la registrazione delle informazioni previste, evitando l'introduzione di entità ridondanti:
 - una modellizzazione corretta della struttura dati assicura non solo un modo più efficiente per la registrazione dei dati
 - ma anche una maggiore facilità d'uso e di analisi dei dati raccolti, soprattutto nel caso di indagini che si ripetono nel tempo
- Spesso accade che chi opera la stesura cartacea del questionario sottovaluti o trascuri completamente la progettazione concettuale:
 - forse a causa di uno scarso interesse o sensibilità o comprensione per gli aspetti informatici del problema, e in particolare della teoria relazionale (database)
 - conoscenze teoriche di base dei sistemi per la gestione informatica dei dati sono necessarie non solo per l'informatizzazione del questionario, ma anche per operare la stesura cartacea.

45

Modello E-R

■ Grafo E-R

- le Entità si rappresentano con i rettangoli
- le Associazioni si rappresentano con le losanghe (rombi)
- gli Attributi si rappresentano con dei pallini



- la freccia indica che gli elementi dell'entità possono entrare nella relazione una o più volte
- le cardinalità possono essere indicate con la notazione (non molto intuitiva):



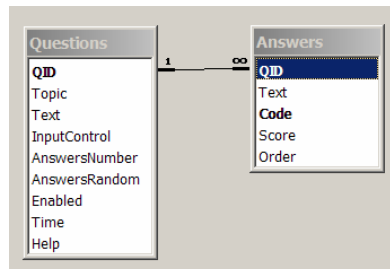
es. una persona può risiedere in una sola città, nella quale possono vivere (da 0 a) n persone

Modello UML

■ Rappresentazione UML

- le entità vengono rappresentate da rettangoli, contenenti :
 - in alto il nome dell'entità
 - all'interno: gli attributi (evidenziando gli attributi che costituiscono la chiave primaria)
- le associazioni sono indicate con una linea di collegamento tra le entità:
 - le cardinalità vengono riportate (più intuitivamente) accanto alla entità, ed in pratica risultano invertite rispetto al diagramma ER
 - notazione molto vicina alla logica del modello relazionale
 es. una domanda può avere n risposte:

- La formalizzazione UML è molto vicina alla progettazione logica, utilizzata nella progettazione dei database relazionali, e adottata da molti DBMS (e CASE), incluso Access



LA PROGETTAZIONE DEL QUESTIONARIO

- Il problema non può essere facilmente ignorato delegando ad un tecnico l'informatizzazione dello strumento, infatti:
 - solo avendo una chiara visione del modello concettuale del questionario è possibile operare una adeguata stesura cartacea della sequenza di domande, che risulti idonea ad essere implementata informaticamente
 - se questo non avviene in fase di stesura cartacea, la successiva fase di informatizzazione richiederà una profonda riprogettazione del questionario
 - spesso il tecnico incaricato non è in grado di operare la necessaria riprogettazione, finendo col limitarsi ad una trasposizione del questionario cartaceo, che risulterà più o meno inadeguata dal punto di vista concettuale
- Il prezzo di una inadeguata progettazione si paga nel momento in cui si vogliono utilizzare i dati rilevati, cioè in fase di analisi, che risulterà oltremodo laboriosa a causa della frammentazione dell'informazione nelle entità ridondanti:
 - domande che si trovano su più percorsi del questionario
 - domande con liste di risposte diverse per categoria di soggetto
- Il procedimento detto di *Normalizzazione* dello schema concettuale del questionario fornisce delle linee guida, che aiutano a tenere sotto controllo la complessità del questionario:
 - valutare l'opportunità di ridisegnare, se necessario, percorsi intricati in una sequenza di domande più semplice e lineare
 - trovare un compromesso tra controllo dell'intervista e ottimizzazione dello schema (*data model*)

IL QUESTIONARIO COMPUTER ASSISTED

- I primi sistemi per realizzare indagini CA nascono negli anni '70 per le indagini telefoniche CATI, campo in cui trovano tuttora il maggior numero di applicazioni.
- I sistemi CATI più sviluppati permettono di informatizzare non solo il questionario ma l'intera organizzazione dell'indagine telefonica: gestione del piano di lavoro giornaliero dei diversi operatori, agenda degli appuntamenti telefonici, scheduling delle chiamate, funzioni di supervisione.
- I sistemi CATI sono stati in seguito utilizzati anche per condurre interviste faccia a faccia CAPI, in cui generalmente un intervistatore raggiunge i soggetti con un computer portatile e registra le risposte per mezzo del programma CAPI.
- Sempre gli stessi sistemi sono recentemente stati utilizzati anche per realizzare indagini CASI, nonostante non si rivelino del tutto idonei a questo scopo, a causa delle particolari problematiche legate all'auto-somministrazione.
- Nel CASI le caratteristiche dell'interfaccia del programma sono particolarmente importanti e risultano critiche per la buona riuscita dell'indagine, infatti in assenza dell'intervistatore, il "questionario" costituisce il solo tramite tra ricercatore e rispondente.
- I moduli di intervista dei sistemi CATI presentano una interfaccia essenzialmente testuale, derivata dalle piattaforme di sviluppo originarie, e ritenute ancor oggi efficaci per l'utilizzo da parte degli intervistatori, ma non sufficientemente *user-friendly* per l'auto-somministrazione

IL QUESTIONARIO COMPUTER ASSISTED

- L'interfaccia dei principali sistemi CATI riflette ancor oggi l'origine, legata alle caratteristiche HW e SW degli anni 70:
 - interfaccia utente di tipo essenzialmente testuale: caratteristica dei programmi che dominavano la scena prima dell'avvento dei sistemi operativi grafici a finestre
 - la compilazione del questionario richiede una certa familiarità con il programma: es. una qualche pratica nell'uso della tastiera (specifiche combinazioni di tasti, ...), il supporto del mouse è secondario
 - solo dopo un periodo di addestramento diventa pratica e molto efficace per l'intervistatore
 - massimo sfruttamento dello spazio a disposizione sul video: schermata fissa, molto compatta, divisa in regioni, una domanda alla volta
- Nell'interfaccia CATI, una regione dello schermo può essere convenientemente dedicata alla visualizzazione di informazioni sull'intervistato fornite nel corso dell'intervista o di interviste precedenti, o note a priori:
 - riduce nell'operatore l'effetto di "isolamento" (o "segmentazione") della domanda dal contesto dell'intervista
 - aiuta l'intervistatore, soprattutto quando il questionario è complesso, a mantenere il quadro della situazione e quindi a gestire meglio l'intervista

IL QUESTIONARIO COMPUTER ASSISTED

- Nell'interfaccia CASI (CSAQ) l'aspetto più importante è la facilità d'uso (*user-friendliness*): il programma deve risultare di immediata comprensibilità per il rispondente, senza richiedere addestramento informatico:
 - evitare videate eccessivamente cariche di informazioni o troppo compatte
 - utilizzare una struttura omogenea per tutte le schermate e convenzioni di formato uniformi per i diversi tipi di oggetti che le compongono
 - ciascuna schermata deve essere formattata in modo che siano immediatamente identificabili le informazioni chiave:
 - il testo della domanda e la lista delle eventuali modalità di risposta precodificate
 - come si fa a inserire le risposte: non dovrebbe essere necessario spiegarlo, si dovrebbe intuire
 - eventuali spiegazioni contestuali in merito alle specifiche domande
 - organizzazione in pagine: riproducendo il layout del questionario cartaceo
- Informazioni ausiliarie e/o istruzioni alla compilazione nel CASI dovrebbero:
 - quando essenziali, essere integrate nel testo della domanda,
 - se invece si prevede che debbano essere utilizzate poco frequentemente, possono essere poste in un help separato, accessibile tramite una apposita icona in fianco al testo della domanda

IL QUESTIONARIO COMPUTER ASSISTED

- Le applicazioni CASI costituiranno probabilmente l'area di maggior sviluppo dei sistemi CA, per la convergenza di diversi fattori:
 - le condizioni di applicabilità della tecnica sono sempre più favorevoli (disponibilità di PC a domicilio, diffusione della banda larga, etc).
 - i costi delle infrastrutture informatiche e di rete sono destinati a ridursi ulteriormente
 - lo sviluppo del software consente di realizzare interfacce grafiche di uso immediato anche per i soggetti non ancora alfabetizzati informaticamente (che sono inoltre in numero sempre inferiore)
- Le tecniche TDE e VRE, eventualmente supportate da ITVD, combinano l'efficacia e la tempestività dell'indagine telefonica con l'economicità tipica dell'autosomministrazione.
- Grande attenzione merita la modalità CAWI (*Computer Assisted Web Interviewing*), cioè CASI on-line via Web:
 - da semplici interfacce grafiche (form statiche) per l'acquisizione e l'invio di dati a un server mediante un bottone (submit), senza reali funzionalità CAI
 - a vere e proprie applicazioni CAI interattive, che utilizzano l'architettura client-server per implementare sul lato server l'elaborazione remota dei controlli di compatibilità in tempo reale durante l'intervista

IL QUESTIONARIO COMPUTER ASSISTED

- Internet potrebbe diventare uno dei più importanti canali di comunicazione per le indagini statistiche:
 - oggi le problematiche collegate alla copertura non consentono di prendere in considerazione questo canale di somministrazione per indagini sulla popolazione, infatti basta considerare che:

$$\bar{y} = \frac{n_o \bar{y}_o + n_m \bar{y}_m}{n} = \frac{n_o \bar{y}_o + (n_m \bar{y}_o - n_m \bar{y}_o) + n_m \bar{y}_m}{n} = \bar{y}_o + \frac{n_m}{n} (\bar{y}_m - \bar{y}_o)$$

$$\Rightarrow \bar{y}_o - \bar{y} = \frac{n_m}{n} (\bar{y}_m - \bar{y}_o)$$

- già oggi risulta tuttavia utilizzabile per indagini su particolari sub-popolazioni, precocemente alfabetizzate informaticamente (studenti, aziende)
- i rispondenti mostrano, almeno per ora, una certa preferenza a rispondere a CAWI rispetto a CASI cartaceo (es. postale)

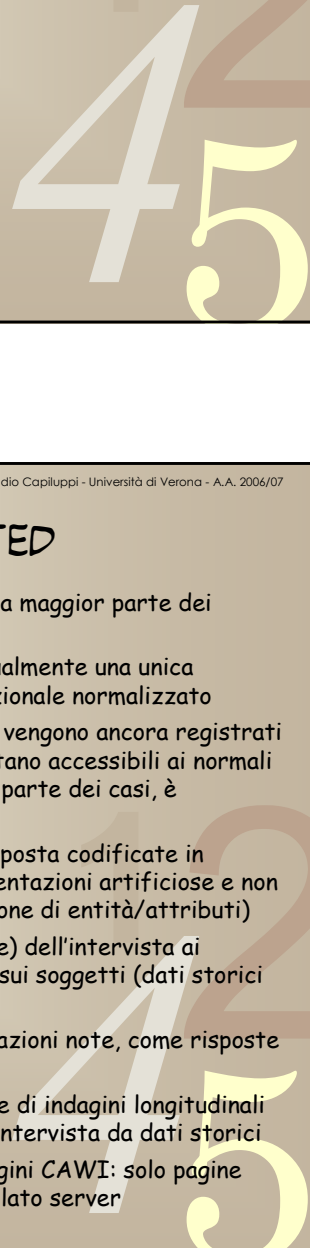
IL QUESTIONARIO COMPUTER ASSISTED

- Caratteristiche sistemi CASI di ultima generazione:
 - costruzione del questionario guidata, interattiva o visuale
 - interfaccia del questionario grafica, di uso immediato anche per chi non ha mai visto un PC
 - molta coreografia (multimediale)
 - in definitiva non molta sostanza: pochi passi avanti significativi per risolvere le problematiche statistiche importanti insite nelle indagini complesse.



IL QUESTIONARIO COMPUTER ASSISTED

- Problematiche non ancora affrontate adeguatamente dalla maggior parte dei sistemi CASI esistenti:
 - struttura logica della data-collection inadeguata: usualmente una unica tabellona bidimensionale, quasi mai un database relazionale normalizzato
 - struttura fisica della data-collection obsoleta: i dati vengono ancora registrati in files binari con formato "proprietario", e non risultano accessibili ai normali software per l'elaborazione statistica (nella maggior parte dei casi, è supportata la sola conversione in ASCII)
 - limitata possibilità di configurazione delle liste di risposta codificate in funzione dei soggetti: si è spesso costretti a implementazioni artificiali e non corrette dal punto di vista concettuale (es. duplicazione di entità/attributi)
 - limitate possibilità di adattamento (personalizzazione) dell'intervista ai soggetti in funzione di informazioni a priori già note sui soggetti (dati storici individuali)
 - controlli di coerenza tra intervista in corso e informazioni note, come risposte fornite in precedenti interviste dagli stessi soggetti
 - in definitiva, supporto inadeguato per la realizzazione di indagini longitudinali prospettive (*panel*) con forte condizionamento dell'intervista da dati storici
 - supporto minimale o nullo per la realizzazione di indagini CAWI: solo pagine html, ma nessun supporto per la programmazione del lato server



ORGANIZZAZIONE DEI DATI STATISTICI

- I dati raccolti durante una indagine possono essere memorizzati su supporto informatico mediante due principali approcci a cui corrispondono strutture dati molto diverse.
- Approccio tradizionale: utilizzato dalla maggior parte dei sistemi CAI, realizzati interamente in un linguaggio di programmazione come Basic, C, Pascal, ...
 - dati registrati su file ASCII: file di testo, delimitati o a campi fissi
 - files dati in formato binario proprietario: record con campi a dimensione fissa
- Approccio relazionale: approccio ben consolidato in campo gestionale, ma non ancora adeguatamente recepito nell'organizzazione dei dati delle indagini statistiche
 - dati strutturati secondo il modello entità-relazione: il file fisico è gestito dall'*engine* del DBMS (*Data Base Management System*) utilizzato
 - dati registrati in un *database server*: un sistema dedicato, specializzato per la gestione centralizzata di dati in rete (es. MySQL, SQL-server, Oracle, ...)
 - utilizzato solo da alcuni (rarissimi) sistemi CAI di ultima generazione, che si appoggiano per la gestione dei dati ad un DBMS relazionale

45

ORGANIZZAZIONE DEI DATI STATISTICI

- Problematiche dell'approccio tradizionale
 - dipendenza dalla struttura fisica (# righe/colonne per i file ASCII , posizione fisica records/campi all'interno di files binari)
 - struttura dati tabellare costituita da una (unica) matrice bidimensionale => ridondanza: duplicazione entità ed informazioni
 - rappresentazione inadeguata dei dati (es. risposte multiple) => subottimalità e artificiosità del modello
 - struttura record gerarchica a dimensione variabile (sub-records) => complessità utilizzo e decodifica
 - accesso sequenziale: operazioni di selezione e ricerca poco efficienti => inefficienza
- Vantaggi dell'approccio relazionale
 - progettazione concettuale e struttura logica indipendente dalla struttura fisica,
 - implementazione fisica completamente a carico del DBMS
 - eliminazione duplicazioni e ridondanze mediante processo di *Normalizzazione*
 - flessibilità: possibilità di riorganizzare i dati nella forma più utile in funzione del tipo di analisi o di interrogazione, mediante query SQL
 - accesso diretto: indicizzazione campi chiave automatica

45