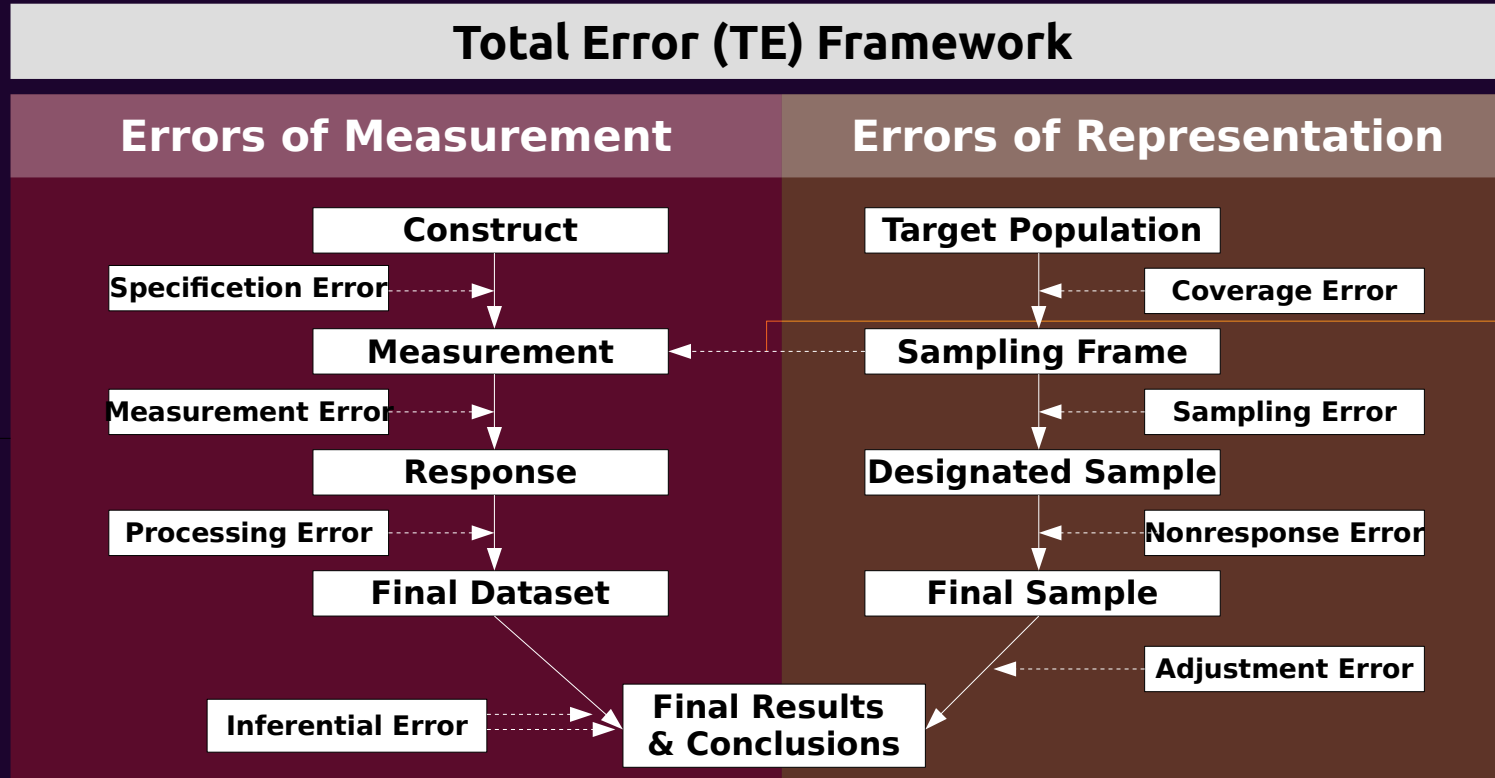


Indagini Statistiche ...

in pratica



Il **Sampling Frame** influenza il modo attraverso cui vengono raccolti i dati

L'errore

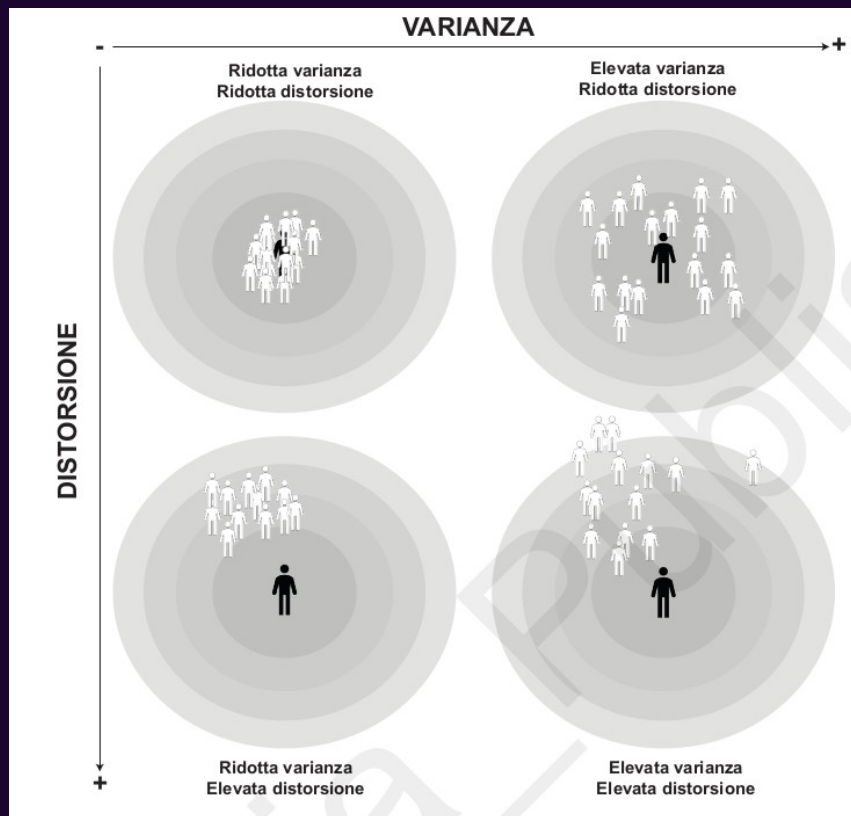
Errore Totale dell'Indagine

=

Varianza

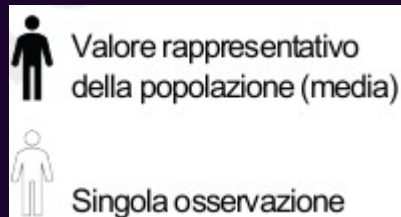
+

Distorsione



Errori casuali

Errori sistematici



Coverage Error

Le liste a disposizione

Elenco telefonico:

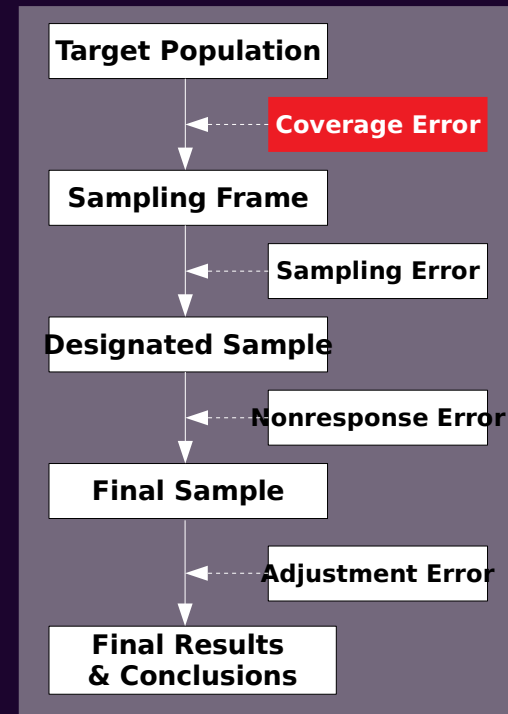
- ▶ 51,3% delle famiglie ha la linea fissa, di cui il 66,1% ha il numero nell'elenco telefonico
- ▶ → 34% della popolazione è rintracciabile nell'elenco

Numeri mobili generati casualmente:

- ▶ 95,9% delle famiglie ha il cellulare

Panel online commerciale:

- ▶ 76,1% ha accesso a internet
- ▶ Tutti i panel italiani max 50.000 complete



Frame a disposizione

CATI

Popolazione

- ▶ Elenco telefonico
- ▶ Lista fornita dal cliente

Aziende

- ▶ Pagine Gialle
- ▶ Acquisto DB
- ▶ Lista fornita dal cliente

CAMI

Popolazione

- ▶ RDD
- ▶ Lista fornita dal cliente

CAWI

Panel online commerciali

- ▶ Liste di persone che rispondono a questionari in cambio di piccoli premi

Liste di email

- ▶ Liste di clienti o persone che hanno aderito a servizi
- ▶ Liste di aziende provenienti da DB amministrativi

Social Media

- ▶ Facebook, Instagram o Tik Tok

River Sampling

- ▶ Questionari proposti su siti, attraverso banner o in App
- ▶ Post acquisto / contatto

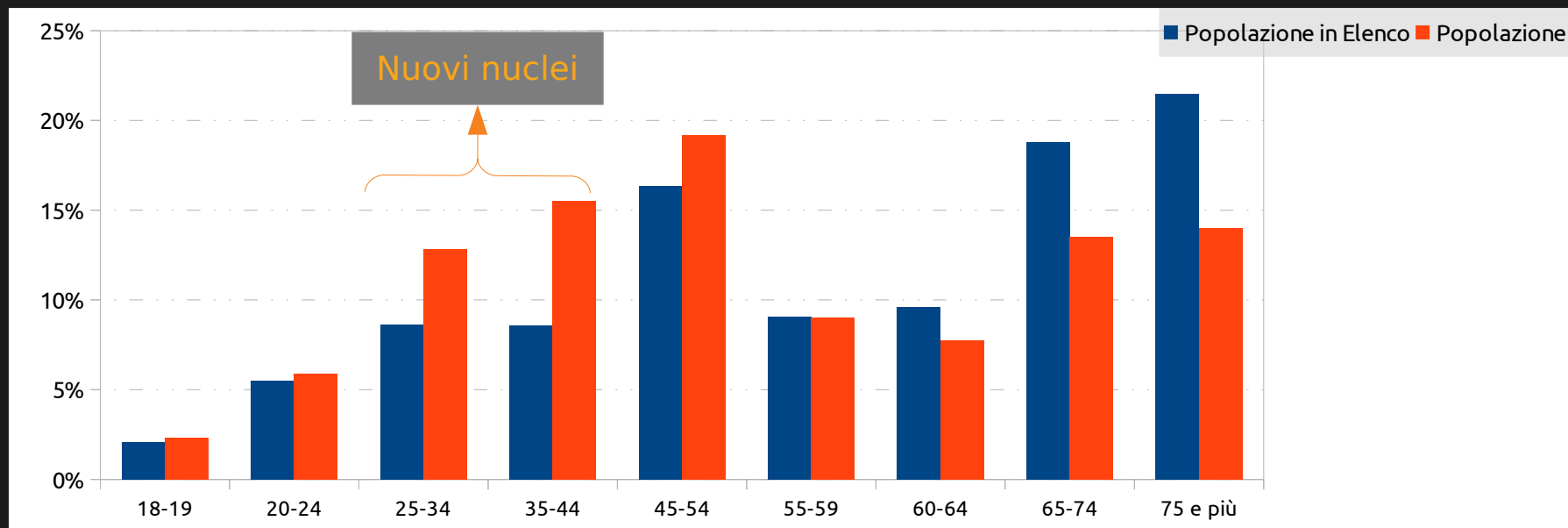
Face to Face Online ➡ Reclutamento via telefono o via web

Popolazione: Elenco Telefonico

Possesso linea fissa: 13.086.000 famiglie (51.3% delle famiglie)

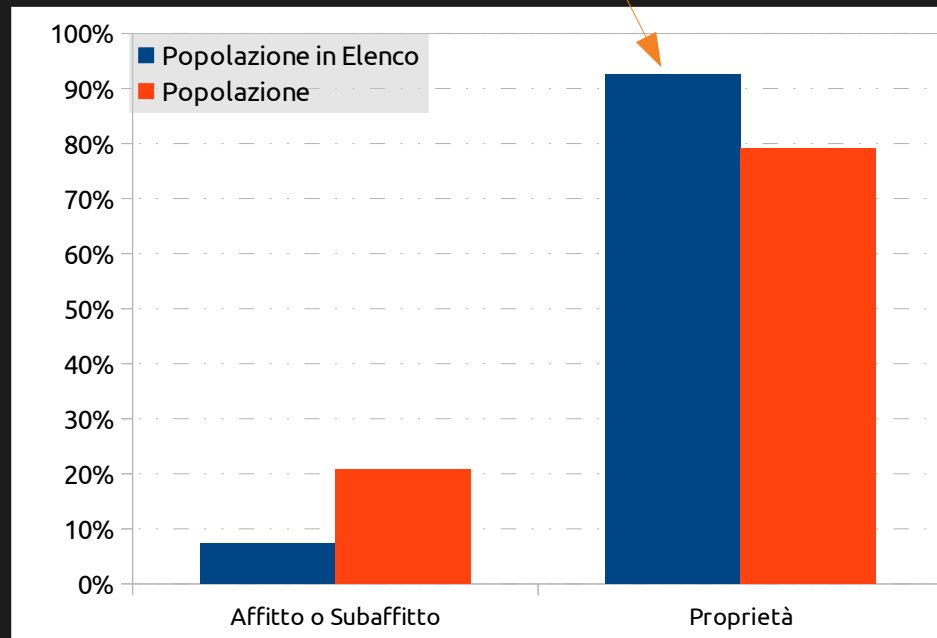
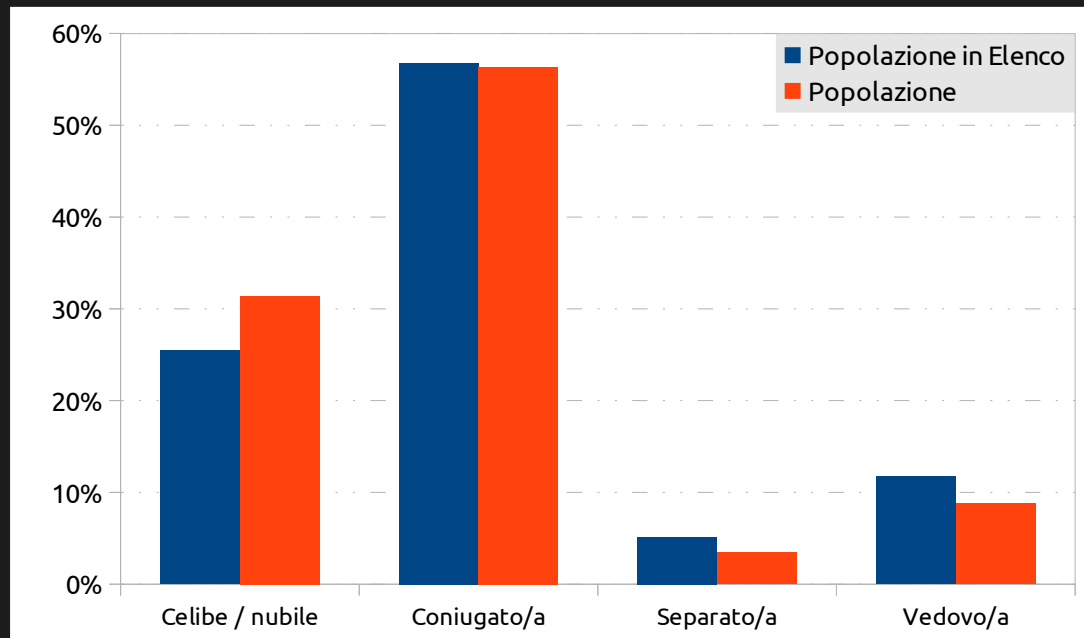
Numero disponibile nell'elenco telefonico: 61,1%

34% della popolazione è rintracciabile sull'elenco telefonico



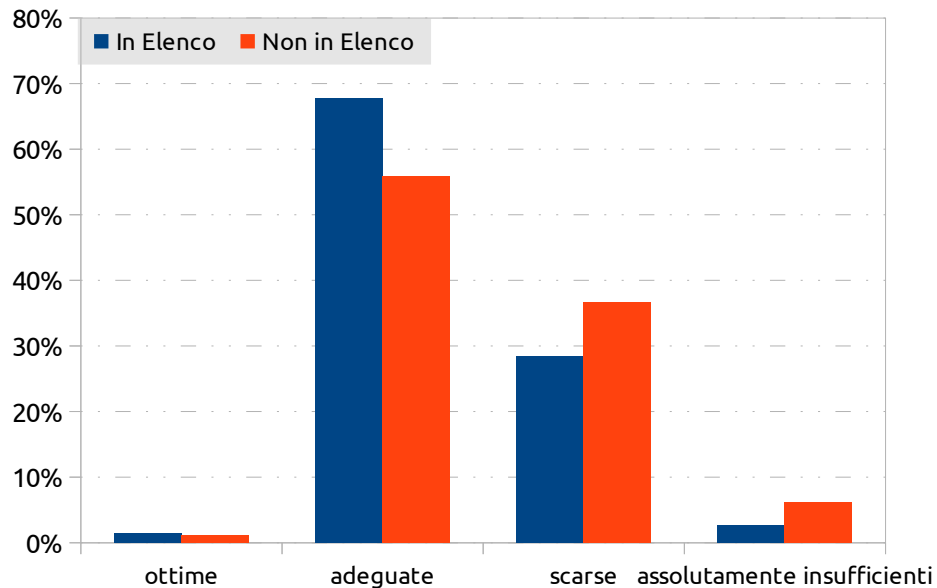
Popolazione: Elenco Telefonico

La quasi totalità di chi è in elenco ha casa di proprietà



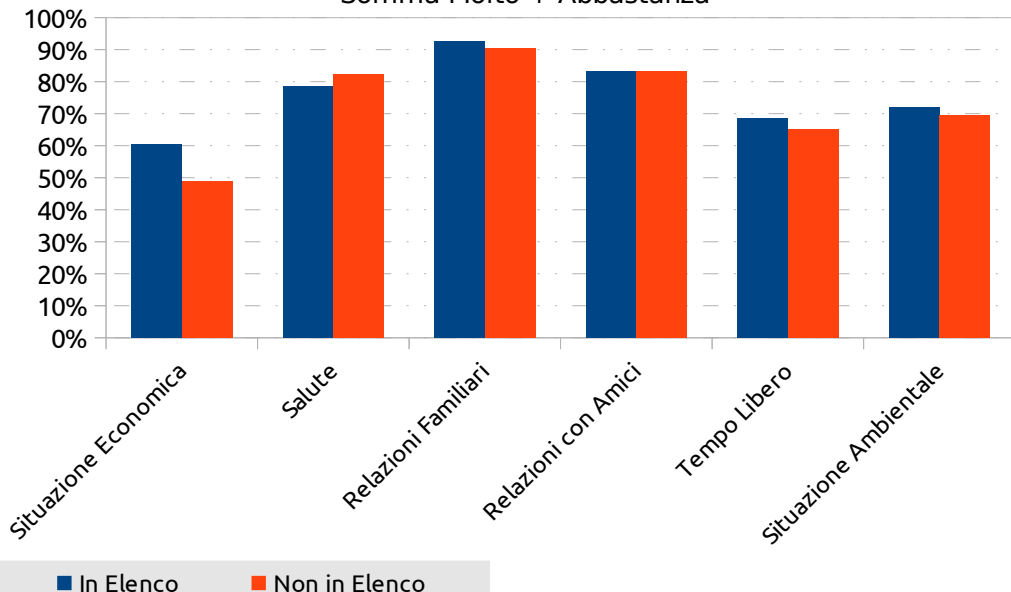
Diverse Condizioni Economiche

Condizione Economica della Famiglia



Soddisfazione per

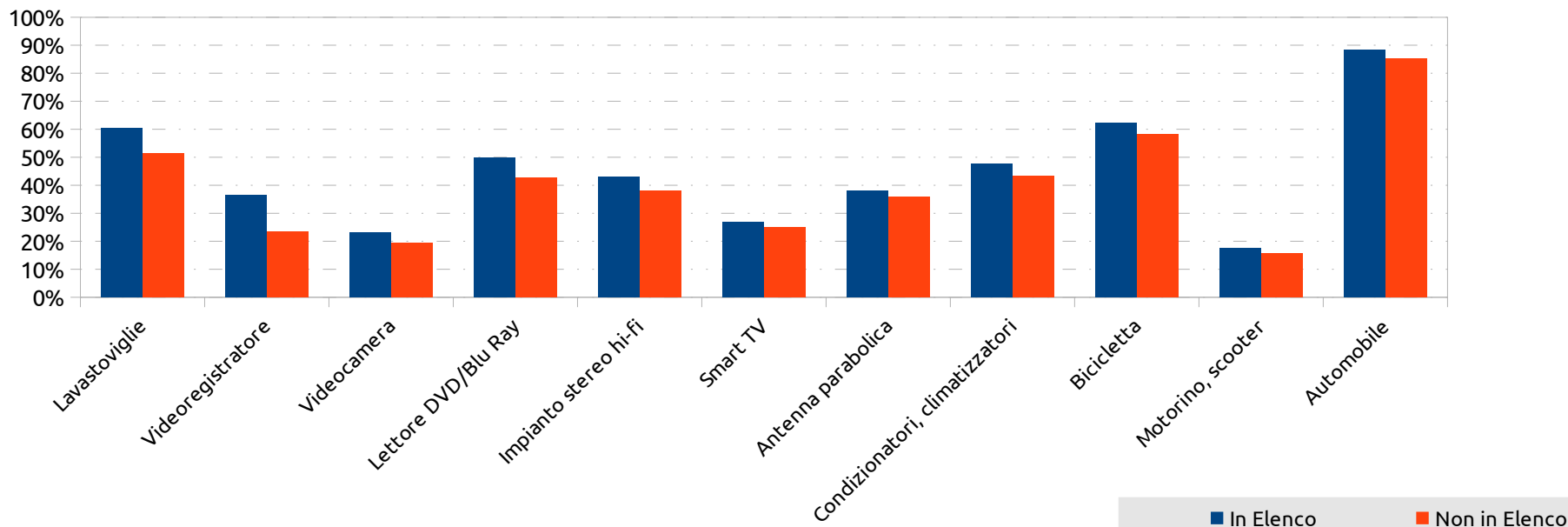
Somma Molto + Abbastanza



Chi è in elenco mediamente vive in una condizione economica migliore

Diverse Condizioni Economiche

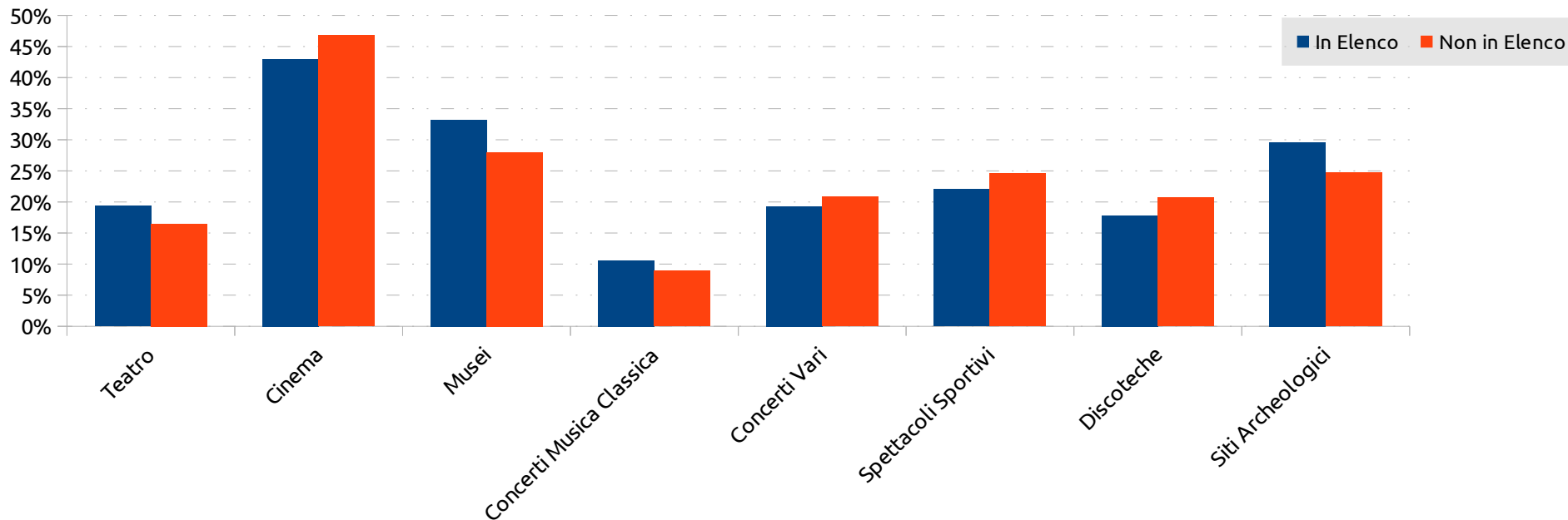
La famiglia Possiede ...



Le persone in elenco hanno in generale una maggiore disponibilità di beni

Diversi modi di passare il tempo

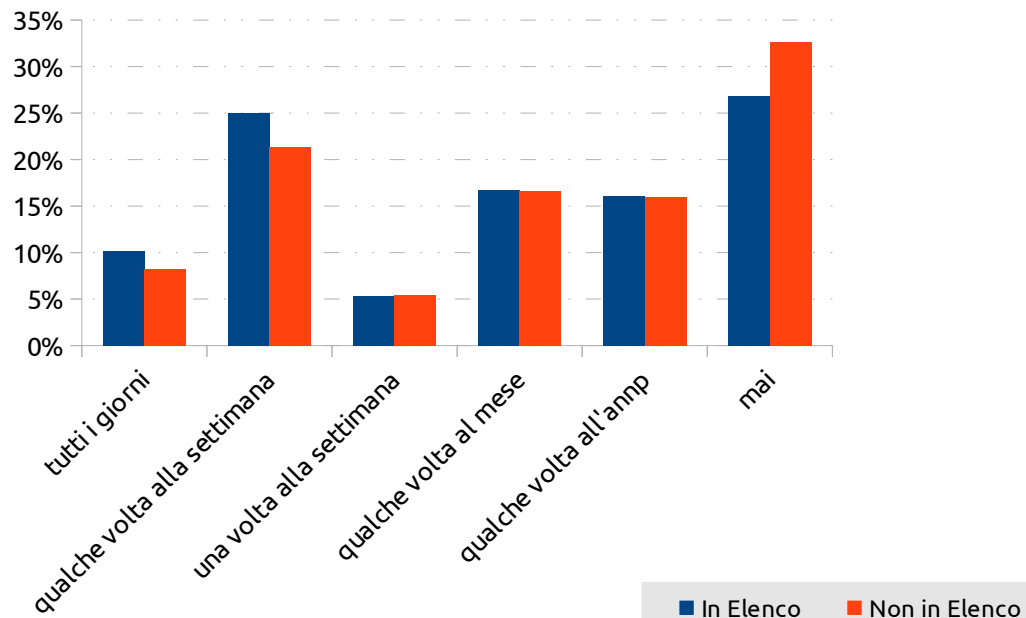
Negli ultimi 12 mesi ha frequentato ...



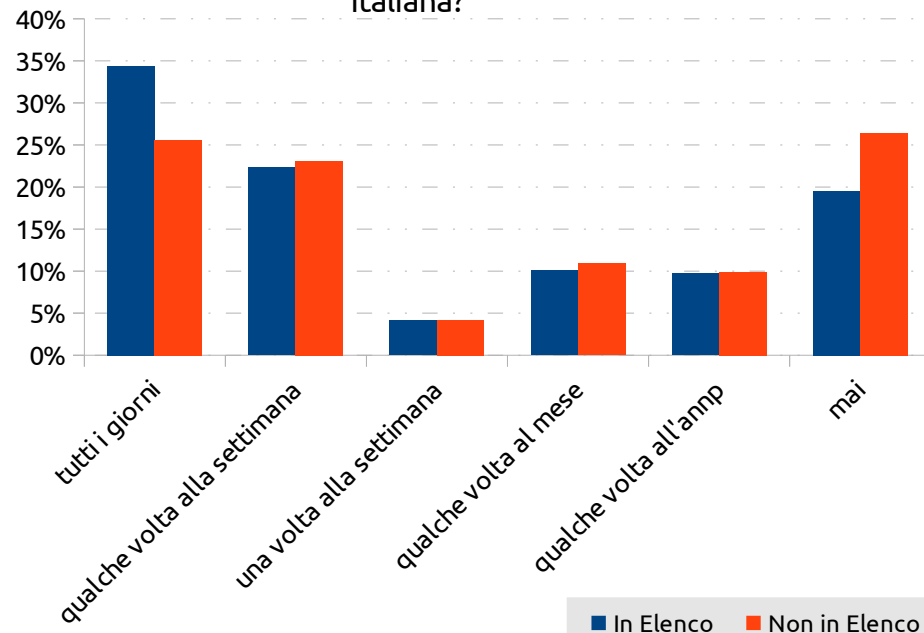
**Chi è in elenco è solitamente
più “acculturato”**

Diverso Interesse per la Politica

Con che frequenza le capita di parlare di politica?



Con che frequenza si informa dei fatti della politica italiana?

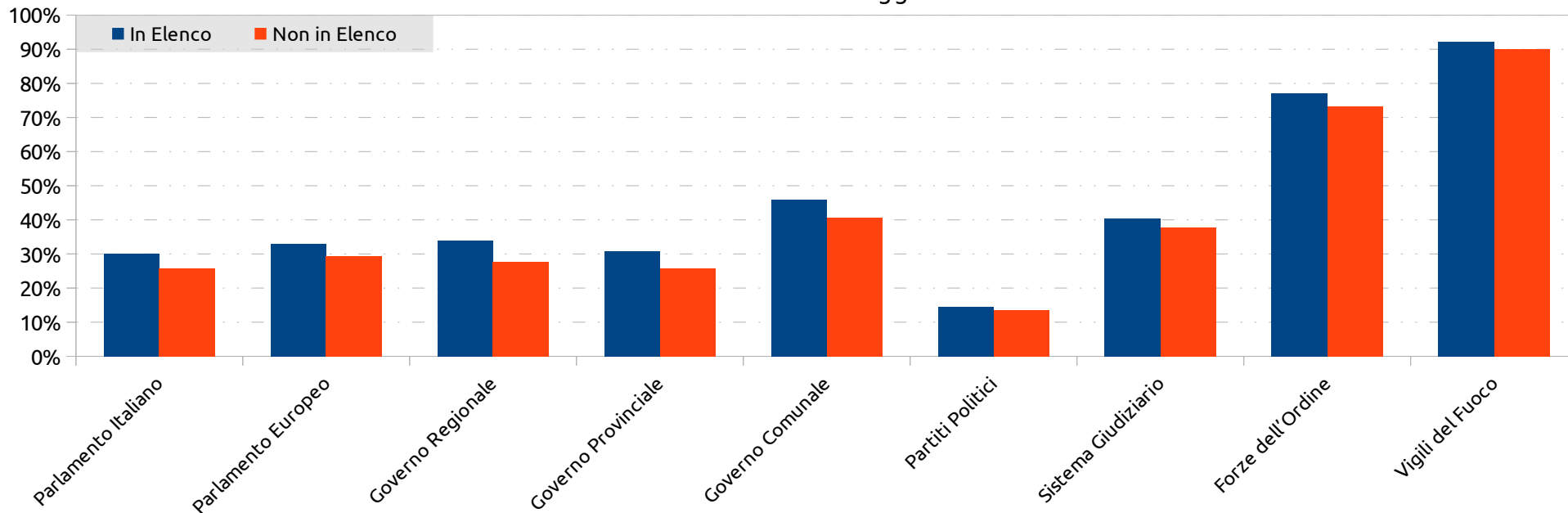


Chi è in elenco si interessa di più di politica

Diverse Opinioni

Quanta fiducia ha ...

Percentuale voti maggiori di 5

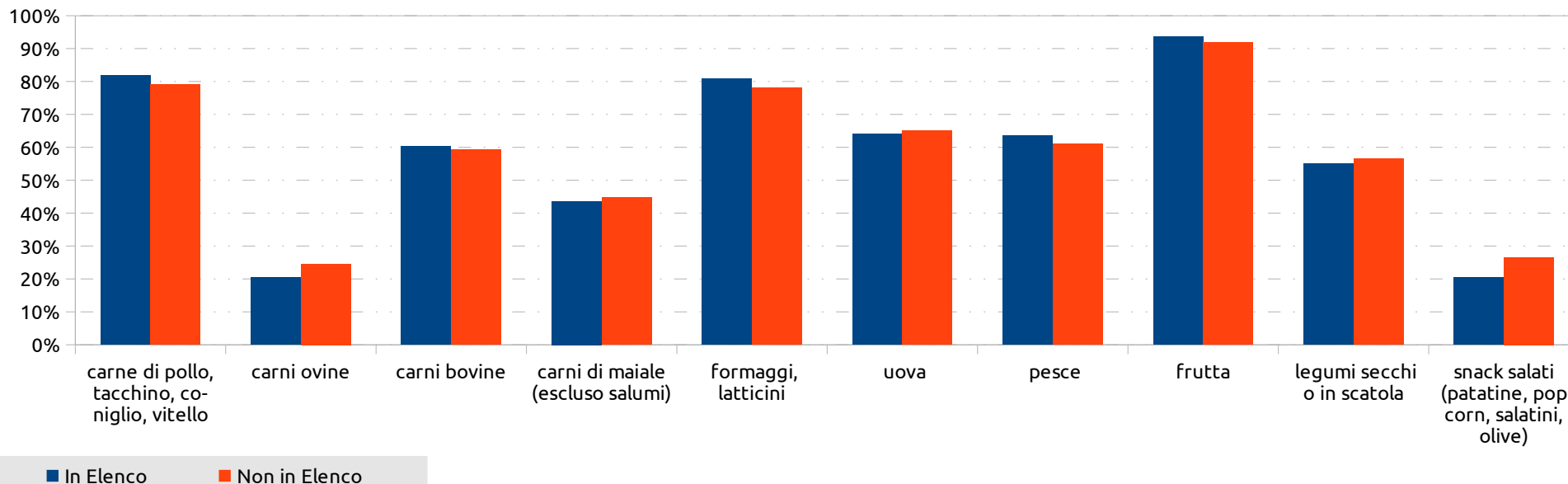


Le persone in elenco hanno in generale una maggiore fiducia nelle istituzioni

Diverse Abitudini di Consumo Alimentare

Frequenza consumo

Almeno una volta al giorno



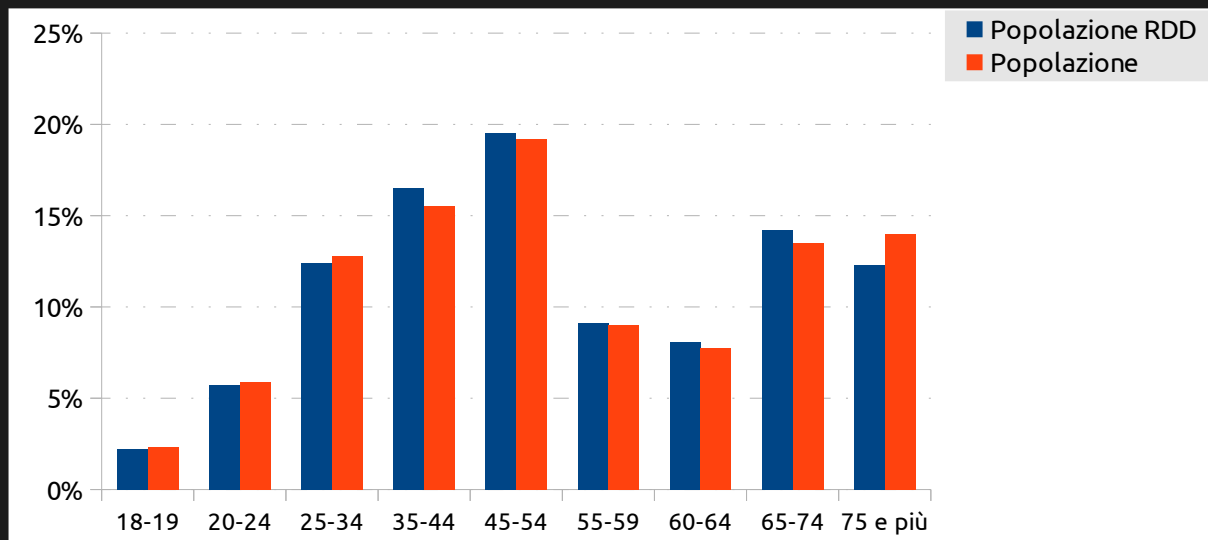
Le persone in elenco hanno un'alimentazione più 'sana'

Popolazione: RDD

Numeri generati casualmente partendo da radici

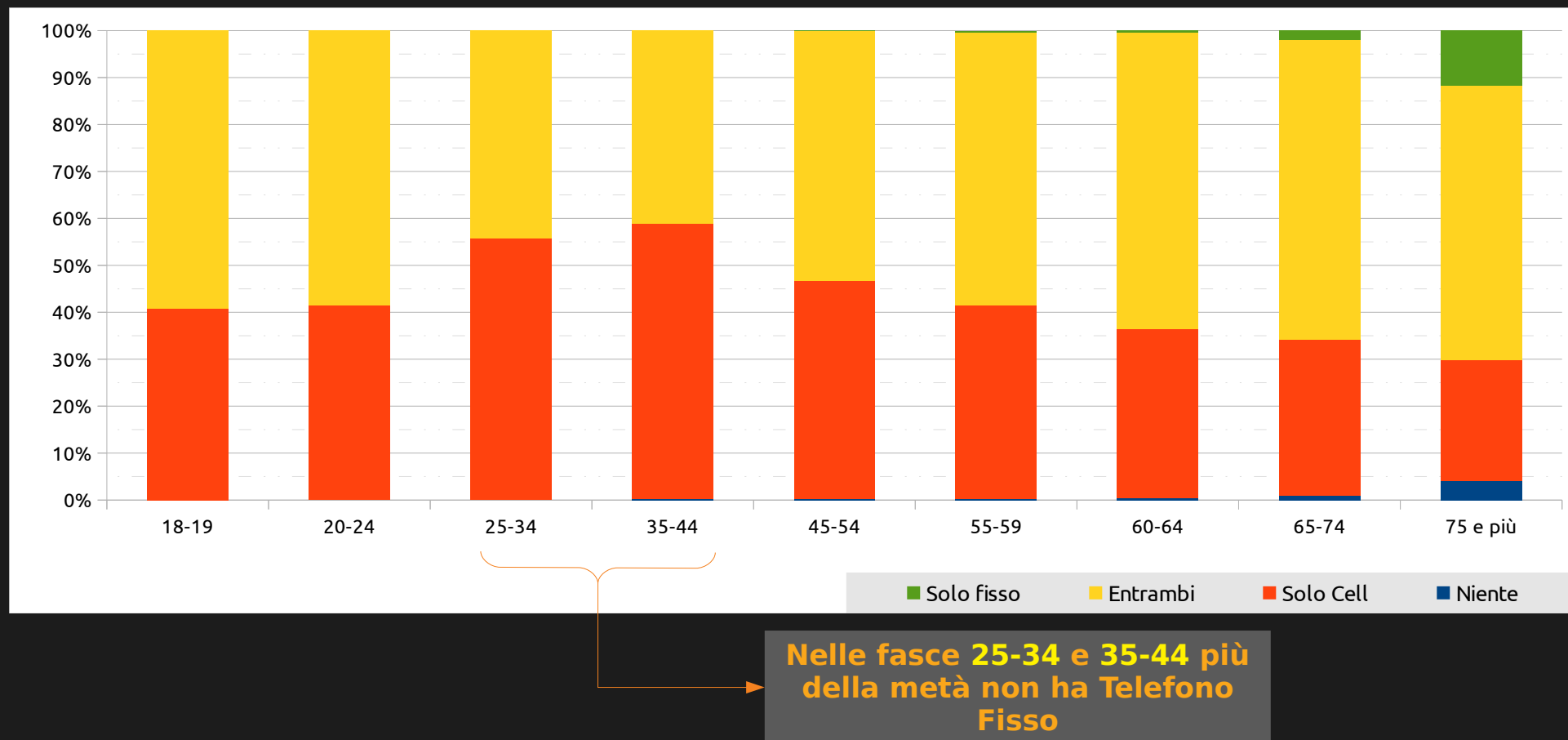
Georeferenziazione fino a livello comunale

Possesso telefono mobile: 24,474,000 famiglie (95,9% delle famiglie)

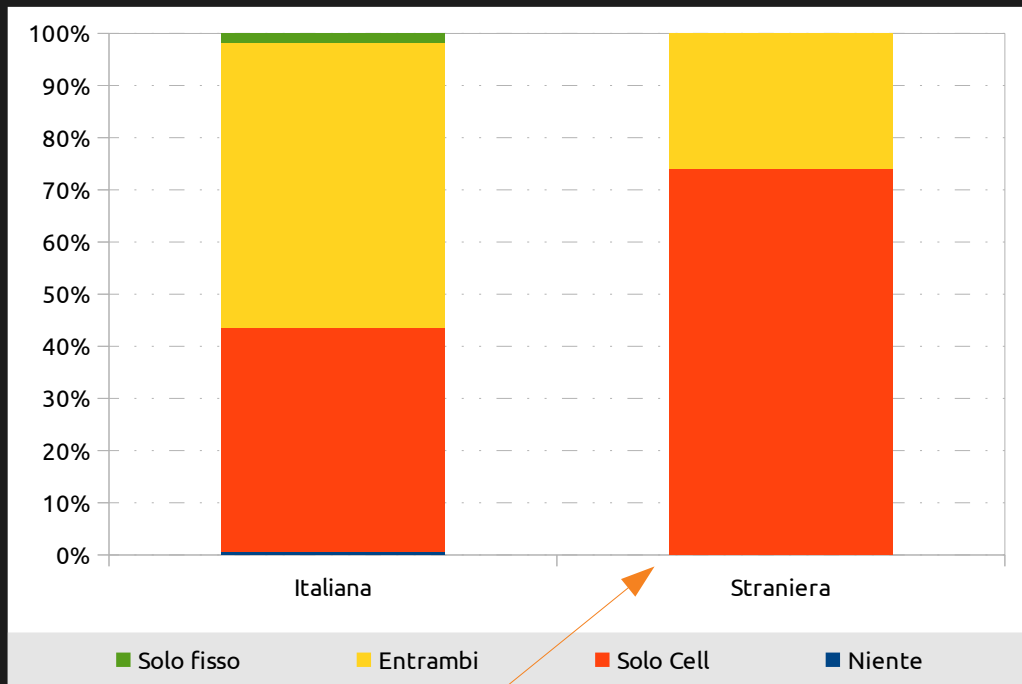


**Buona
distribuzione per
classi d'età**

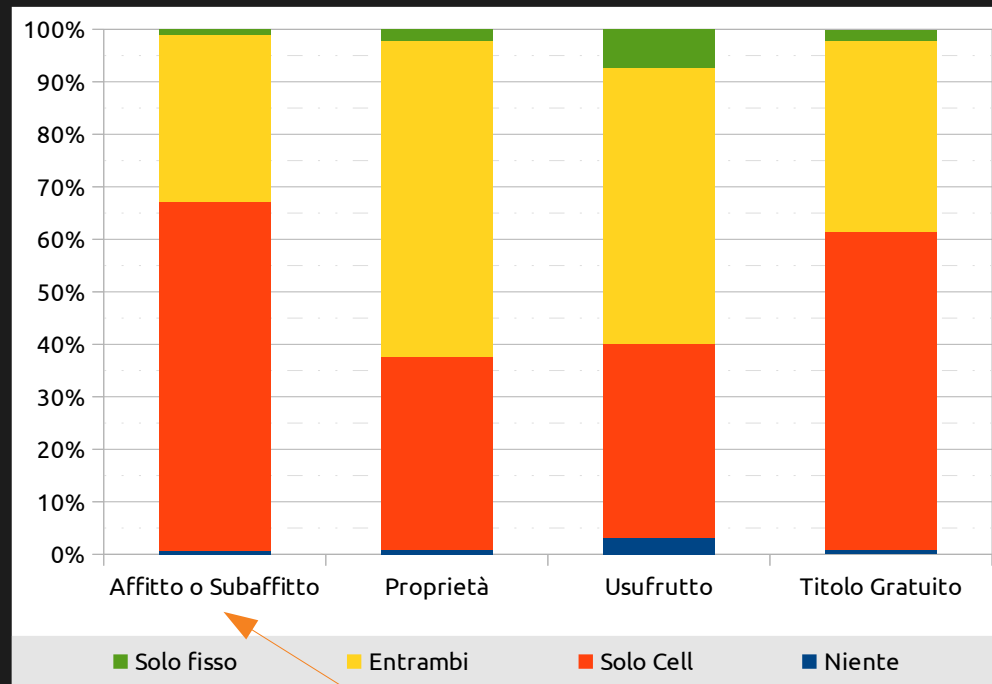
Quindi? Dove si trova chi?



Quindi? Dove si trova chi?

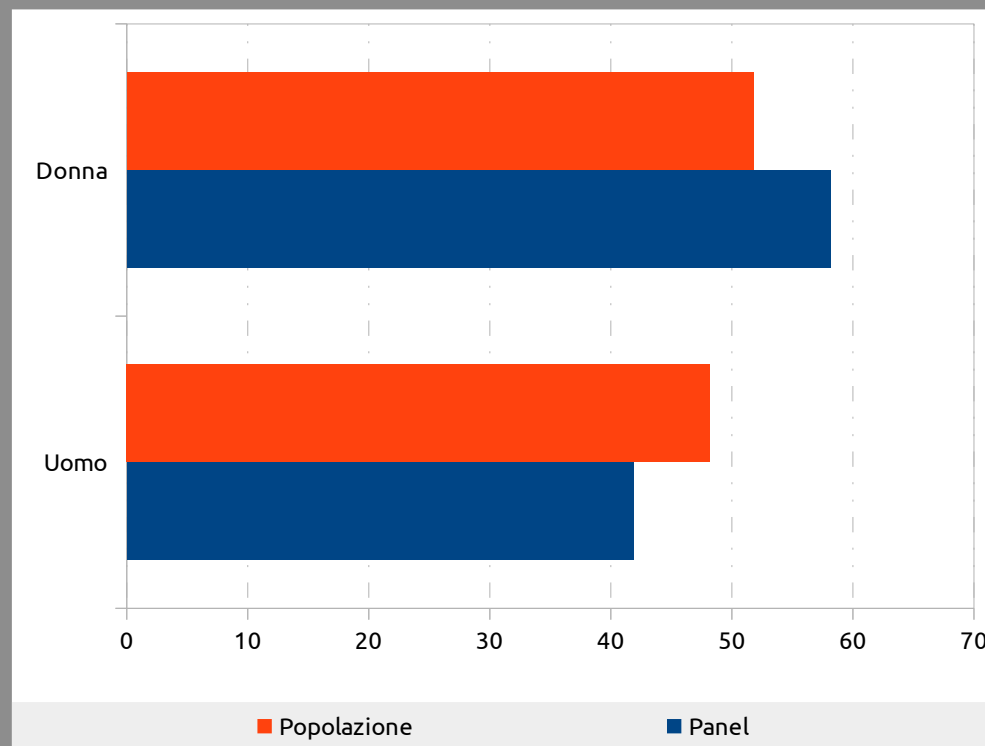
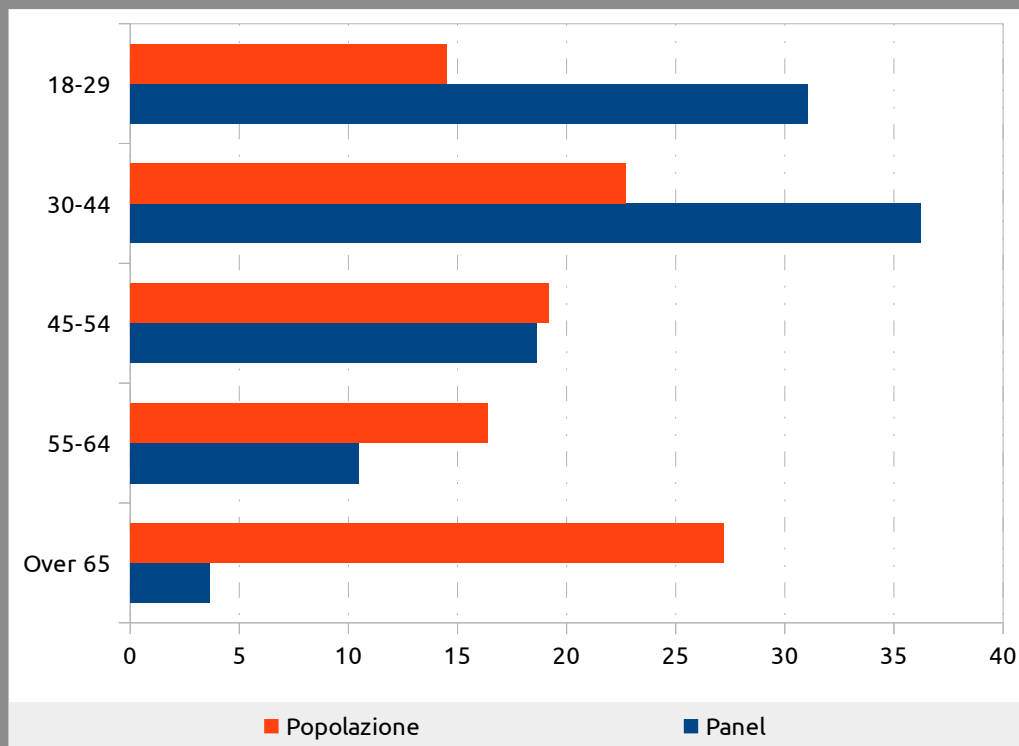


Più del 70% degli stranieri residenti in Italia non hanno il telefono fisso

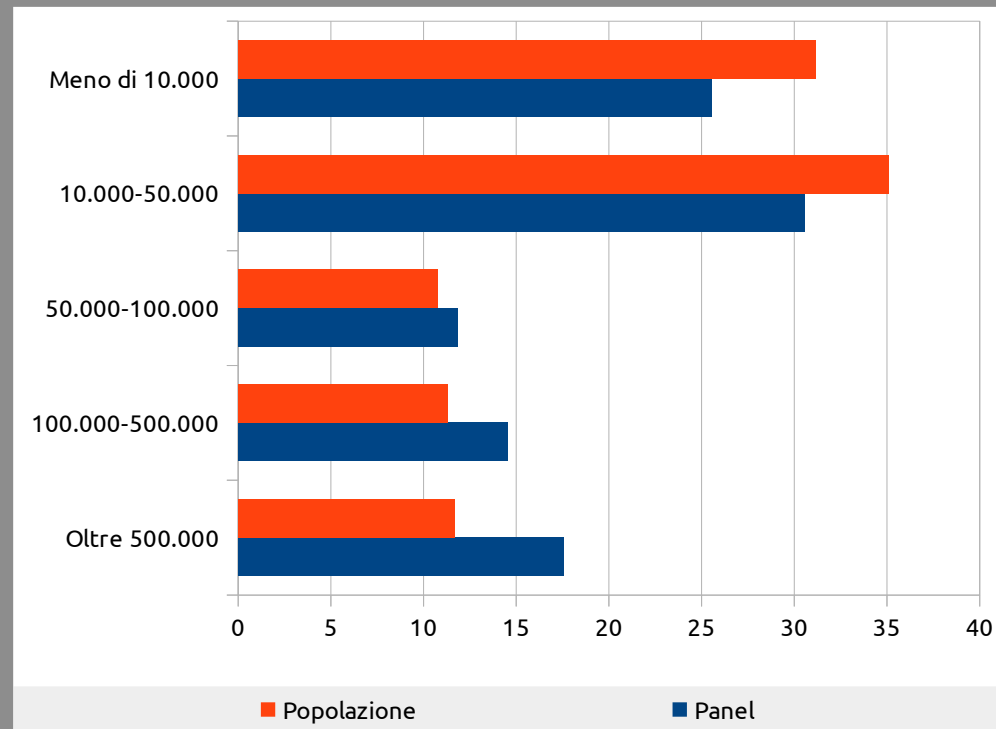
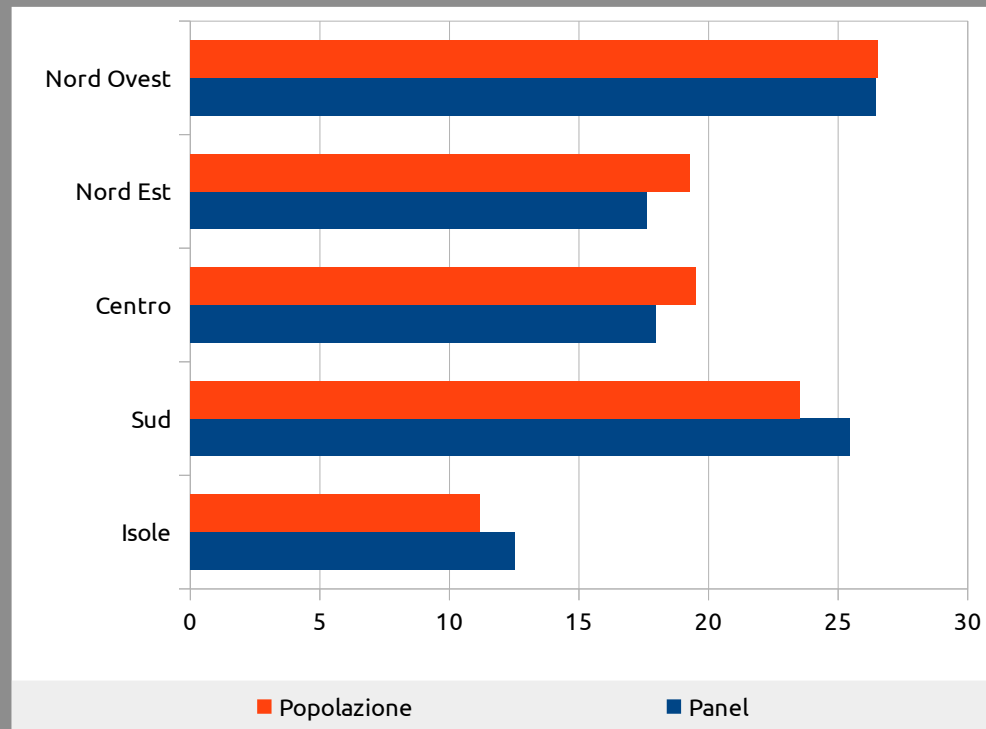


Tra gli affittuari oltre il 60% non ha telefono fisso

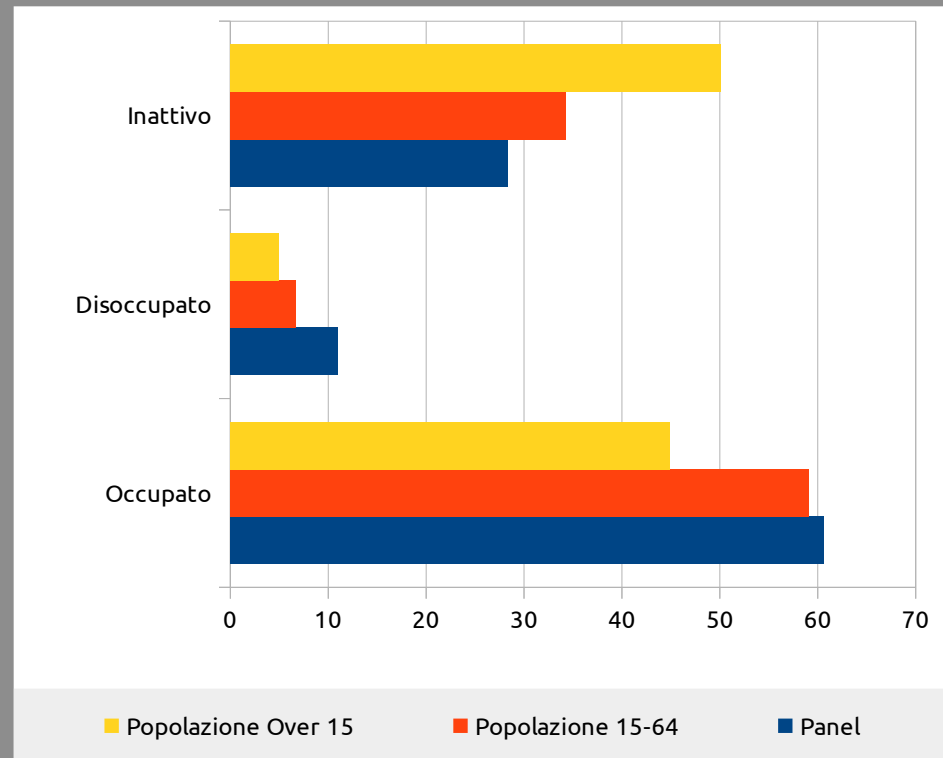
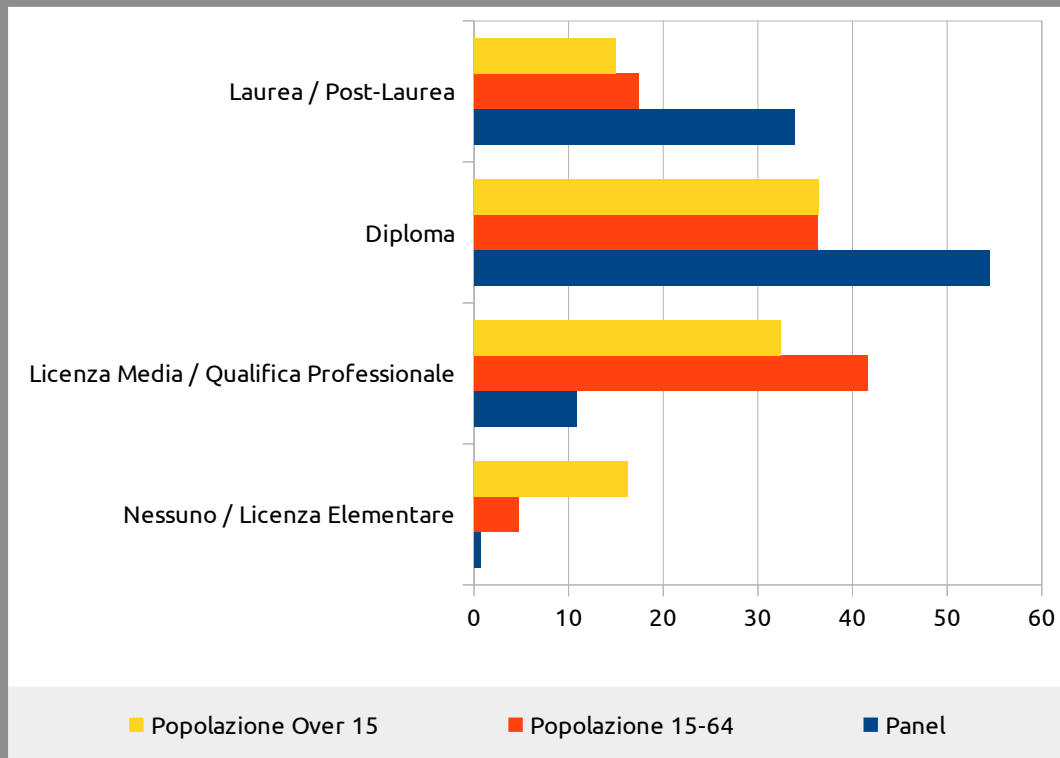
Panel Online Commerciali



Panel Online Commerciali

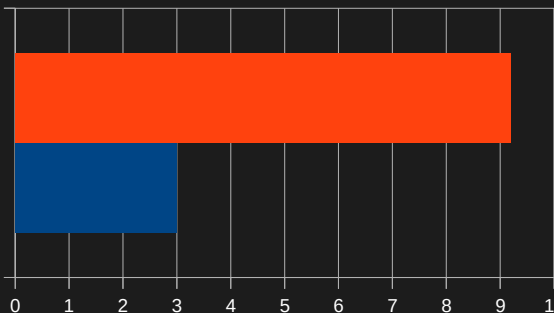


Panel Online Commerciali

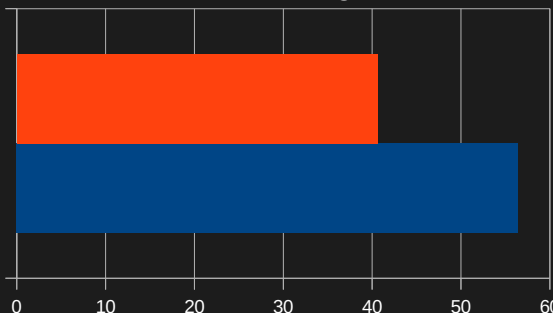


Panel Online Commerciali

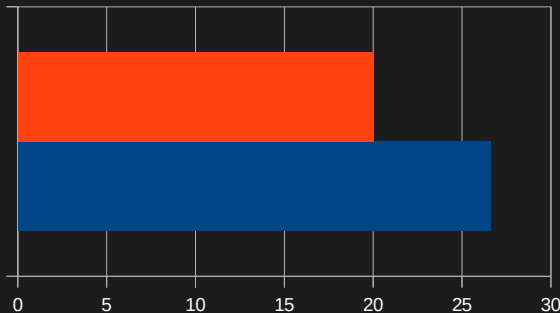
Non guardo la TV



Ascolto la radio tutti i giorni

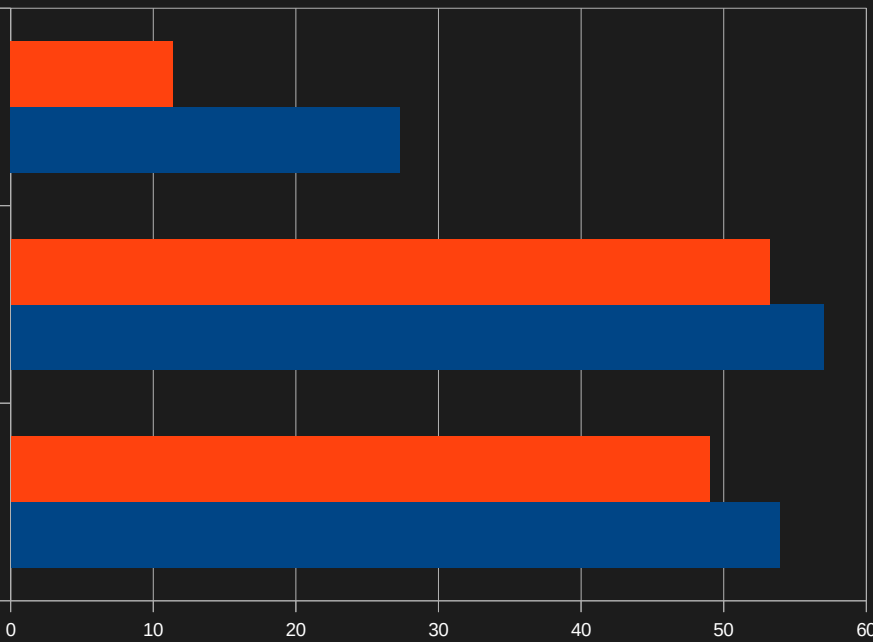


Fumo

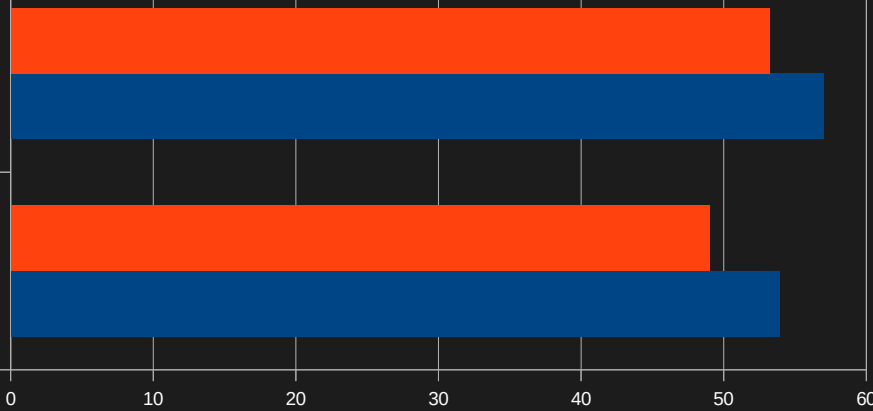


■ Panel Online
■ Multiscopo

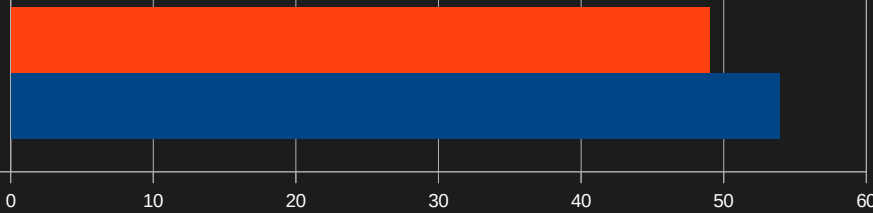
Consumo superalcolici



Consumo vino



Consumo birra (nazionale)



Dati pesati per Genere-età
e Titolo di Studio

Differenze significative al
95%

Sampling Error

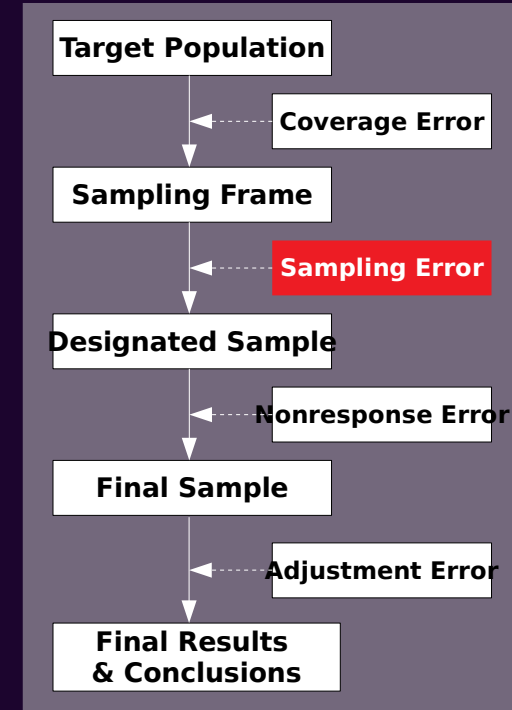
I tipi di campionamento

Casuale Semplice: Le Unità Statistiche vengono estratte casualmente dal Sampling Frame

Stratificato: Le Unità Statistiche vengono estratte in base all'appartenenza a degli strati definiti a priori (**es:** zona di residenza)

- ▶ **Proporzionale:** la numerosità di ogni strato è calcolata in maniera proporzionale rispetto all'universo
- ▶ **Ragionato:** la numerosità degli strati è definita in modo 'ragionato' per ottenere numerosità minime in ogni strato

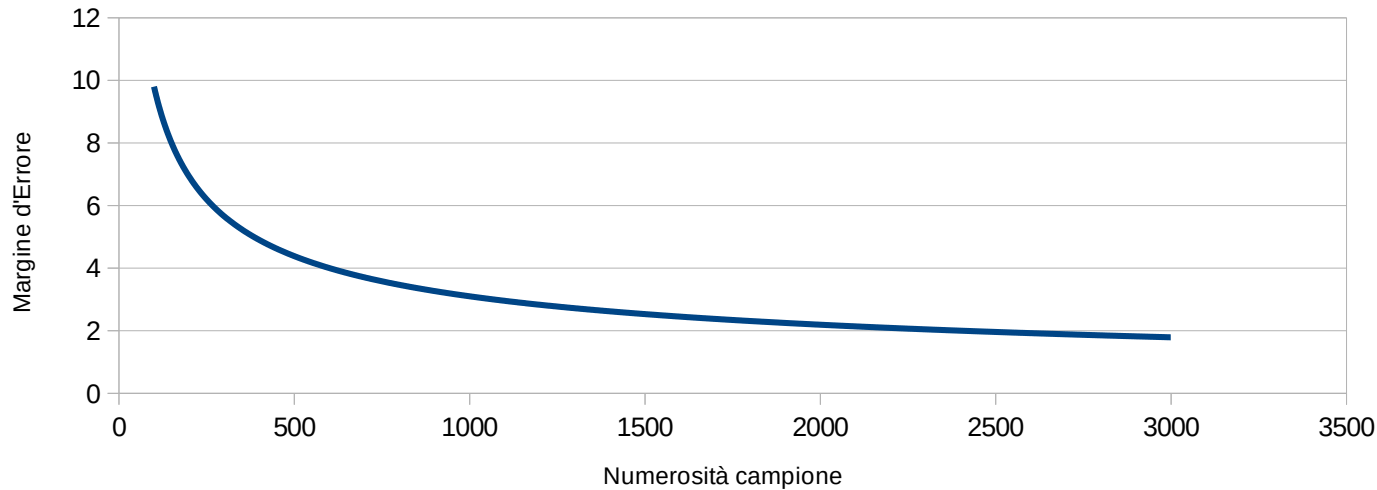
Per Quote: Si definiscono delle quote (**es:** genere-età) che vanno a chiudersi mano a mano durante l'indagine



Sampling Error

Il Margine d'errore

Andamento del Margine d'Errore al crescere della Numerosità Campionaria



Media in popolazione = media campione $\pm$ margine di errore

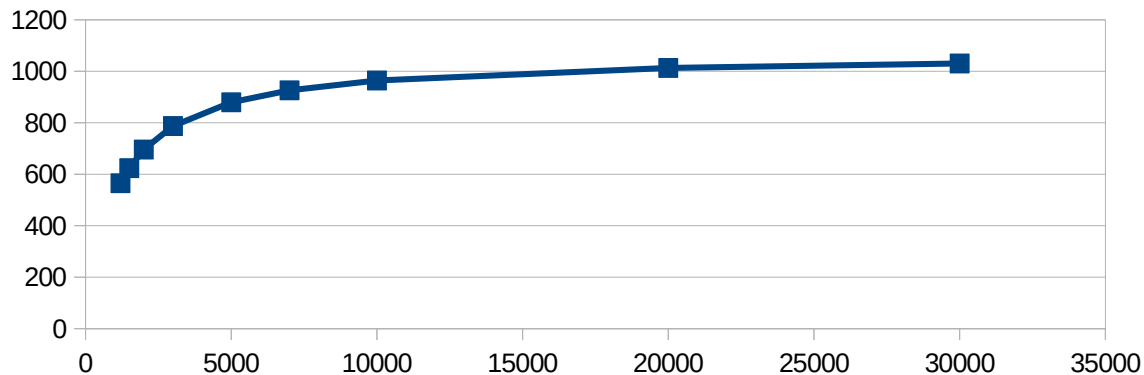
► **Es:** media campione = 15
Margine di errore = 2
 $13 \leq \text{Media in popolazione} \leq 17$

$$Z * \sqrt{\frac{\sigma^2}{n} * \frac{N-n}{N-1}}$$

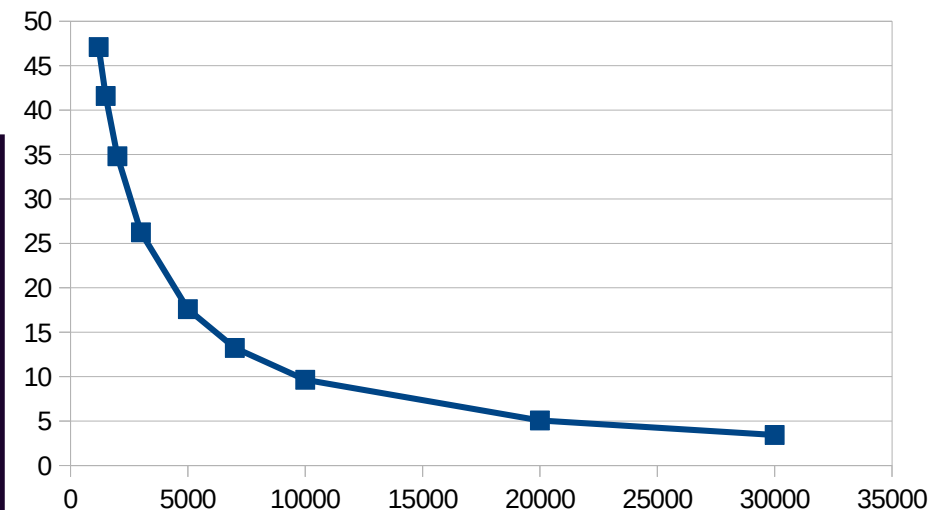
$Z = 1.96$ livello di
confidenza al 95%

$\sigma^2 = p * (1-p) \rightarrow$
ipotesi conservativa:
 $0.5 * 0.5 = 0.025$

Sampling Error



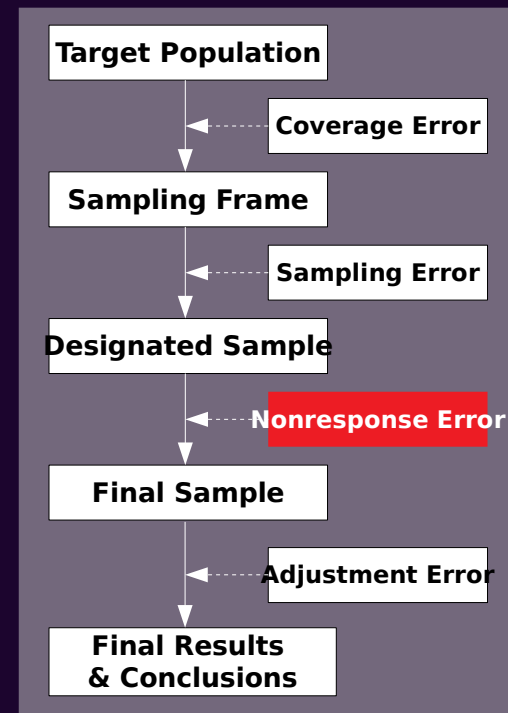
Numerosità del campione per ottenere un margine d'errore del 3% al variare della dimensione della popolazione obiettivo



Percentuale della popolazione che va campionata per ottenere un margine del 3% al variare della dimensione della popolazione obiettivo

Non Response Error

- ▶ Il **Campione selezionato** non è sempre quello poi effettivamente ottenuto a causa della **Non Risposta**
- ▶ Se la **Non Risposta** è in qualche modo collegata alla **Variabile d'Interesse**, allora si ha l' **ERRORE di NON RISPOSTA**
- ▶ L'**Errore di Non Risposta** è tanto più grande quanto più la **Non Risposta** è collegata alla **Variabile d'Interesse** e tanto più è alto il tasso di **Non Risposta**



Un po' teoria ...

Leverage-Saliency Theory (Groves, Singer, Corning, 2000):

La partecipazione all'indagine e la qualità delle risposte fornite sono funzioni di fattori d'attrazione e fattori dissuasivi

Fattori d'attrazione

- ▶ Argomento
- ▶ Sponsorship
- ▶ Incentivo (**Callegaro**: "The odds of response were at least doubled using monetary incentives")

Fattori dissuasivi

- ▶ Complessità, Sforzo Cognitivo
- ▶ Lunghezza



E la pratica ...

Nelle indagini che svolgiamo possiamo agire solo su alcuni di questi fattori:

Fattori d'attrazione

- ▶ Incentivo

Fattori dissuasivi

- ▶ Complessità, Sforzo Cognitivo
- ▶ Lunghezza

La PARTECIPAZIONE e la QUALITA' dei dati possono essere influenzati da uno sbilanciamento di questi fattori:

- ▶ Incentivo troppo basso per lo sforzo richiesto
- ▶ Lunghezza eccessiva che causa noia



Adjustment Error

Pesatura

Sample weights:

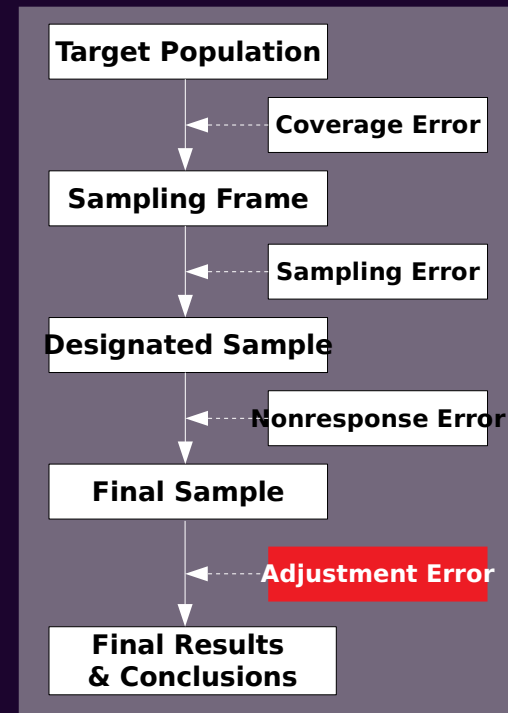
Basati sulle probabilità di selezione, si utilizzano quando non tutte le unità hanno la stessa probabilità di essere estratte
(**es:** campioni stratificati non proporzionali)

Nonresponse weights:

Servono per correggere i tassi di risposta differenti in diversi sottogruppi, si basano sulle variabili utilizzate per la definizione delle quote
(**es:** genere – età)

Post – Stratification weights:

Utilizzati per riportare ai totali di popolazione la distribuzione di variabili che non erano state usate nella stratificazione iniziale ma che si pensa siano collegate con la variabile d'interesse
(**es:** titolo di studio)



Conclusioni

Errori & Soluzioni (o almeno ci si prova)

Coverage error

- Utilizzo più frame
- Domande di screening a inizio questionario

Sampling Error

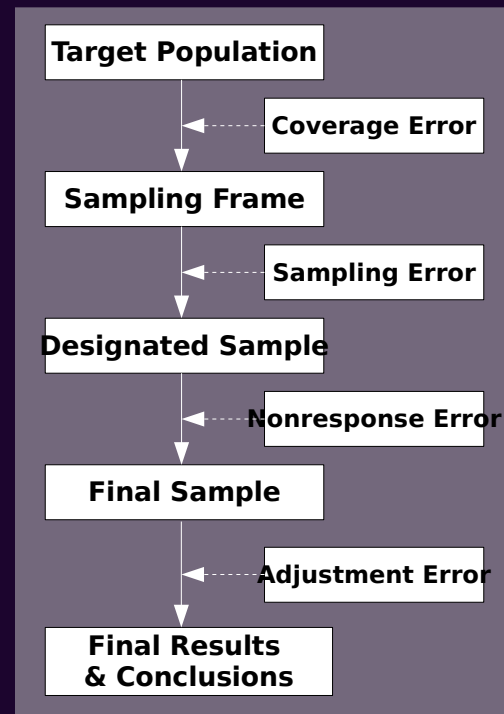
- Utilizzo campioni casuali
- Calcolo numerosità ottimale

NonResponse Error

- Procedure per ridurre le non risposte
- Pesatura dei dati

Adjustment Error

- Calcolo di diversi tipi di pesi
- Valutazione delle stime ottenute



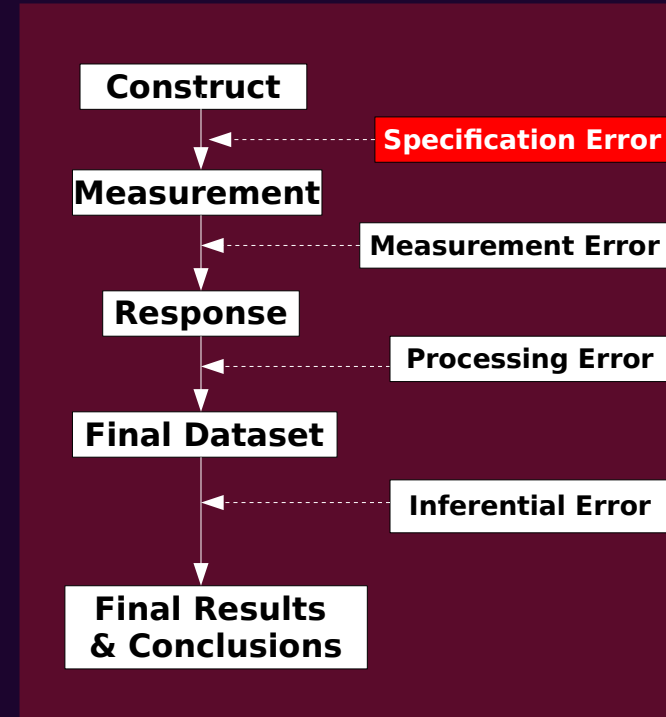
Specification Error

Il costrutto misurato da una domanda e il costrutto che si voleva misurare differiscono

Non tutti i costrutti sono direttamente osservabili, molti sono latenti e composti da più dimensioni

Non includere una dimensione, porta allo Specification Error

Le dimensioni poi vanno tradotte in indicatori



Specification Error

Concetto Generale: Definizione del Fenomeno da indagare

Individuazione delle diverse DIMENSIONI di cui è composto il fenomeno

Concettualizzazione:
Definizione utile in funzione
Della teoria che voglio verificare / indagare

Indicatore: Identificazione delle manifestazioni del fenomeno rilevabili

Definizione delle VARIABILI che saranno rilevate

Categoria: Identificazione dei possibili stati della variabile

Definizione Operativa:
Standardizzazione

Specification Error

Esempi

Concetto

Dimensione

Indicatore

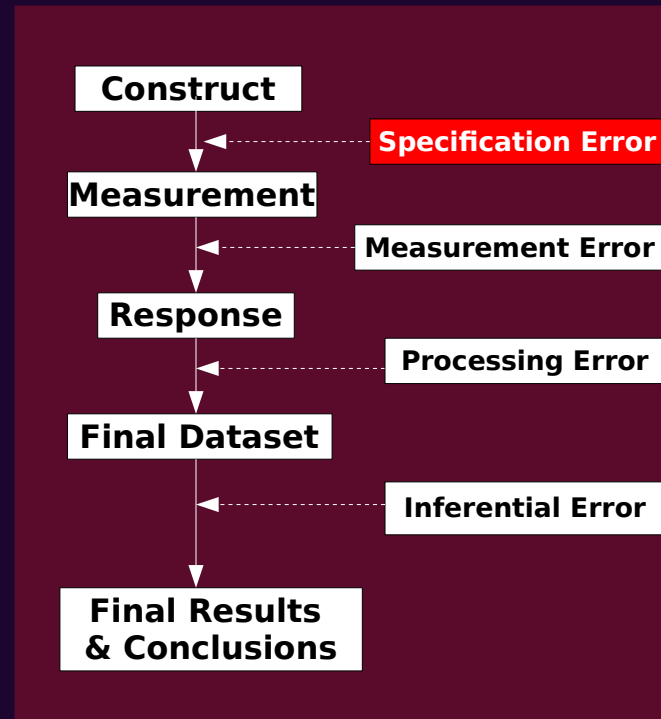
Variabile

Categoria

Livello di educazione scolastica

Titolo di studio più alto

Anni di scuola



Specification Error

Esempi

Concetto

Dimensione

Sottodimensioni

Indicatore

Variabile

Categoria

Estremismo di destra

Attitudini

Comportamenti

Xenofobia, Antisemitismo,
Supporto ai dittatori di destra,
...

Voto, Appartenenza,
Provocazione, Violenza

Per ogni Sottodimensione vanno inseriti uno o più Indicatori

Construct

Measurement

Response

Final Dataset

**Final Results
& Conclusions**

Specification Error

Measurement Error

Processing Error

Inferential Error

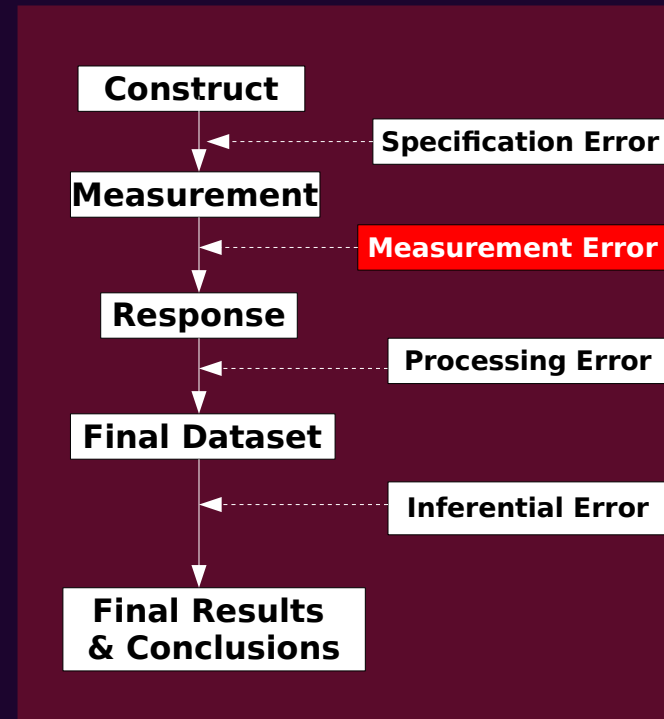
Measurement Error

La risposta raccolta non corrisponde alla realtà, è influenzata da fattori esterni

Può causare errori sistematici ed errori casuali, quindi può introdurre sia distorsione che aumento della varianza

L'errore di misura può essere causato da diversi fattori :

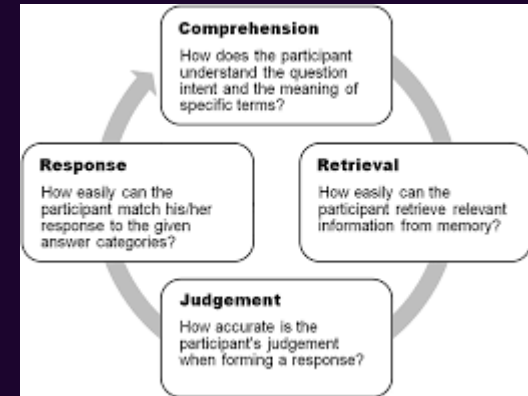
- ▶ Metodo di somministrazione del questionario
- ▶ Questionario
- ▶ Intervistatore
- ▶ Rispondente



Il processo di risposta secondo Tourangeau (2000)

Il processo di risposta è composto da 4 fasi, ciascuna delle quali può causare problemi o bias

1. **Comprensione:** Ascolto, comprensione e interpretazione della domanda
→ *Difficoltà nel testo della domanda, domanda non chiara*
2. **Retrieval:** Recuperare, nella memoria, le informazioni necessarie per rispondere
→ *Richiesta di informazioni difficilmente reperibili (es: quanto hai speso in medicine nello scorso anno?)*
3. **Giudizio:** Integrare le informazioni in un giudizio
→ *Desiderabilità sociale*
4. **Risposta:** Tradurre il giudizio in una risposta (compatibile alla scala proposta) e rispondere alla domanda
→ *Scala non adeguata, Acquiescenza*



Satisficing

Satisficing Debole

Il rispondente passa attraverso i 4 passaggi del processo di risposta in modo SUPERFICIALE

- ▶ Acquiescenza
- ▶ Scelta della prima risposta 'accettabile'
- ▶ Stima o 'tiro a indovinare' nelle domande sui comportamenti

Satisficing Forte

Il rispondente SALTA uno o più passaggi del processo di risposta

- ▶ Ricerca delle parole chiave
- ▶ Risposte casuali
- ▶ Risposte neutrali o 'Non so'

Satisficing ... perché?

Motivazioni

- ▶ Il rispondente è annoiato
- ▶ Le norme sociali (CATI) o il desiderio di guadagnare l'incentivo (CAWI) gli impediscono di interrompere
- ▶ Lo sforzo si è rilevato maggiore del previsto

Fattori

- ▶ Difficoltà delle domande
- ▶ Capacità cognitive del rispondente
- ▶ Motivazioni del rispondente a fornire risposte ottimali

 **Leverage-Saliency Theory**

Desiderabilità Sociale e Sensibilità

Motivazioni

- ▶ Dipende da cultura a cultura
- ▶ Possono essere considerate intrusive (es comportamenti illegali)
- ▶ Possono mettere in pericolo l'immagine che si vuole dare di sé (es comportamenti 'sconvenienti')
- ▶ Direttamente proporzionale al livello di presenza dell'intervistatore

Risultato

- ▶ Sovrastima comportamento 'desiderabile'

Esempi

- ▶ Domanda sensibile: 'Ha mai fatto uso di droghe?'
- ▶ Domanda affetta da desiderabilità sociale: 'Ha mai tradito?'