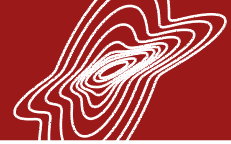


# METODI STATISTICI PER LA BIOINGEGNERIA

A.A. 2024-2025

Prof. Alessandra Bertoldo

Ing. Mattia De Francisci, Ing. Claudia Tarricone




- ESERCIZIO 1

## Cos'è ELSA?

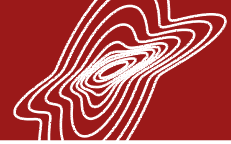
ELSA è uno studio epidemiologico multidisciplinare su un campione rappresentativo di persone residenti in Inghilterra con età maggiore di 50 anni.

Il suo scopo è capire le complesse dinamiche dell'invecchiamento che intercorrono tra le relazioni familiari, affettive, economiche, sociali, biologiche e come si legano alla salute e al benessere delle persone.



## Measures

<b>Demographic data</b> Including Ethnicity, Marital status, and Education	<b>Physical Health</b> Including Mobility, Pain, and Disability
<b>Income and assets</b> Including Earnings, Pensions, and Housing wealth	<b>Behavioural health</b> Including Smoking, Physical Activity, and Sleep
<b>Employment</b> Including Employment status, Job details, and Reasons for retirement	<b>Mental health</b> Including Psychiatric problems, General Health Questionnaire, and CES-D depression scale
<b>Psychosocial factors</b> Including Control, Demand, Effort-reward balance	<b>Psychological and social well-being</b> Including Quality of life (CASP-19) and Personality
<b>Consumption</b> Including Transfers, Housing, and Expenditures	<b>Cognitive Function</b> Including Memory, Literacy, and Fluid intelligence
<b>Expectations</b> Including on Mortality, Employment, and Finance	<b>Psychosocial factors</b> Including Control, and Effort-reward balance
<b>Social and civic participation</b> Including Transport, Social support, and Loneliness	<b>Physical examination and performance</b> Including Blood pressure, Grip strength, and Balance



- ESERCIZIO 1

Cosa trovate nel file `data_ELSA.mat` nella cartella `data`:

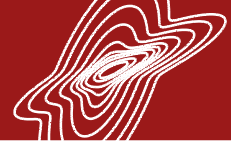
**elsa**: matrice di dimensioni 1000x16

Ogni riga corrisponde ad un diverso soggetto e contiene le realizzazioni delle variabili per quel soggetto.

Ogni colonna corrisponde ad una diversa variabile clinica (e.g., age, BMI (body mass index), heart rate...)

**elsa\_labels**: array di celle di dimensioni 1x16

Ogni elemento corrisponde ad una diversa etichetta delle variabili cliniche.



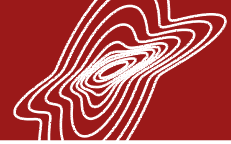
### ESERCIZIO 1 – una prima occhiata ai dati

- Il codice va inserito in uno script matlab (es. lab3\_20241023.m)
- Controllare la bontà del dataset ELSA, verificando:
  - Se le variabili presenti sono continue e se possono assumere valori negativi
  - l'eventuale presenza di valori non numerici **NaN**, derivanti da dati non disponibili o da errori di trascrizione. Quanti sono i valori **NaN** nell'intera tabella? (**isnan**)
  - l'eventuale presenza di valori non numerici **Inf**, derivanti da dati non disponibili o da errori di trascrizione. Quanti sono i valori **Inf** nell'intera tabella? (**isinf**)
  - Verificare l'eventuale presenza di valori negativi, derivanti da dati non disponibili o da errori di trascrizione. Quanti sono i valori negativi nell'intera tabella?
  - Cosa accade se si applica l'operatore media/varianza in presenza di questi valori?



## ESERCIZIO 2 - Pulizia dei dati

- Pulizia a livello di variabili (colonne della matrice):
  - Calcolare la percentuale di **NaN** per ciascuna variabile
  - Qual è la percentuale di **NaN** per la variabile **age**? E per **gluc**? Farne il display sulla Command Window (comando **disp**)
- Se la percentuale di **NaN** è superiore al 20% eliminiamo la variabile dalla matrice. La matrice originale avrà quindi ora un numero di colonne minore rispetto alla partenza.
- Salvare nella variabile **elsa\_bk** un backup della variabile data (per BONUS)



- Pulizia a livello di soggetto (righe della matrice):
  - Selezionare ed eliminare i soggetti che hanno almeno un **NaN**
  - Selezionare ed eliminare i soggetti che hanno almeno un **Inf**
  - Selezionare ed eliminare i soggetti che hanno valori negativi (se la variabile assume solo valori non negativi)
  
  - Quante variabili/soggetti sono rimasti nella matrice? (Risposta: 826x6)
  - Cosa accade se si applica ora l'operatore media/mediana/varianza?
- [BONUS] In **data\_bk** rimpiazzare i valori **NaN** con il valore mediano di ciascuna variabile e sostituire i valori negativi con valori nulli (zero).