

# Analisi dei dati

Domenico De Stefano

a.a. 2025/2026

# Indice

- 1 Dettagli sul corso
- 2 Alcuni appunti sul report statistico

# Orario

- Mercoledì 15-18 Aula 2C (edificio H3)
- Giovedì 16-18 Aula A (edificio A)
- esercitazioni e parte pratica (da definire): ipotesi Venerdì pomeriggio

La parte pratica inizierà indicativamente a fine Marzo, verrà data comunicazione via **Moodle**

# Programma del corso

Due parti: **indagini statistiche e dati in campo sociale (social surveys)** e **analisi dei dati (data analysis/data mining)** e **modelli statistici**

## 1 Indagini statistiche

- ▶ **Indagini statistiche e dati statistici in campo sociale** I dati. Dati e scienze sociali. Le indagine statistiche. Esempi di indagine statistiche a livello nazionale (in particolare indagini ISTAT). Popolazione (Universo) e campione. Il campionamento (concetti base). Studi osservazionali e sperimentali. Nuove fonti dati (es. Social networks). Tecniche di rilevazione dati. Nozioni di campionamento statistico e alcune tra le principali indagini campionarie in Italia ed Europa. Matrice dati e variabili. Indicatori statistici.

# Programma del corso (2)

## 2 Analisi dei dati

### ► Tecniche di analisi dei dati e alcuni modelli statistici.

Introduzione all'analisi dei dati da indagine. Introduzione all'analisi delle reti sociali. Tecniche di analisi dei dati multidimensionali. Elementi di analisi fattoriale (analisi in componenti principali e analisi delle corrispondenze). Analisi dei gruppi. Introduzione ai modelli statistici con particolare attenzione al modello di regressione lineare multiplo e modello logistico.

### ► Parte pratica Il software *R* (<https://cran.r-project.org/>)

# Libri di testo e riferimento

- Indagini statistiche
  - ▶ Materiale liberamente scaricabile dal sito dell'ISTAT (url fornito dal docente via *moodle2* (*moodle2.units.it*))
- Analisi dei dati e modelli statistici
  - ▶ Capitoli da: De Lillo, Argentin, Lucchini, Sarti, Terraneo. **Analisi multivariata per le scienze sociali**. Pearson Italia
- Per entrambe le parti: Slides, appunti dalle lezioni e materiale didattico integrativo fornito dal docente → *moodle2* (*moodle2.units.it*)
- Parte pratica
  - ▶ Stefano M. Iacus e Guido Masarotto. Laboratorio di statistica con R, 2/Ed. McGraw-Hill Education

# Esame

La valutazione finale consisterà in 2 parti: una prova scritta ed un lavoro di gruppo (prova pratica)

- La **prova scritta** della durata massima di 100 minuti, vale 20 punti e consisterà in dieci domande vero/falso e a risposta multipla (0.5 punti per ciascuna risposta esatta e che copriranno maggiormente gli aspetti teorici del corso) ed almeno due esercizi su indagini statistiche e modelli che gli studenti dovranno risolvere, illustrando il procedimento adottato.
- La prova pratica consisterà nell'applicazione del software R per l'analisi di un caso studio (concordato con il docente) i cui risultati verranno presentati mediante slides (massimo 13 punti). Si tratterà di un **lavoro di gruppo** (gruppi di 2-3 persone)

# Utilizzo di moodle

- messaggistica da parte del docente
- quiz a risposta multipla da svolgersi dopo la fine su blocchi di argomenti del corso (introduzione alle indagini, modelli statistici, ecc.)
  - ▶ I quiz sono da intendersi esclusivamente per autovalutazione e per comprendere se ci sono criticità di apprendimento
  - ▶ Il consiglio è di svolgerli sempre anche se non avete avuto tempo di approfondire gli argomenti!
- il forum
  - ▶ Sarà possibile aprire thread all'interno del forum su vari argomenti sia da parte del docente che da parte degli studenti

# Indice

- 1 Dettagli sul corso
- 2 Alcuni appunti sul report statistico

# Parte pratica: report statistico

- Un report statistico viene redatto con lo scopo di informare i lettori sui risultati che possono provenire dalla ricerca sperimentale, dai dati di interviste e questionari o da altre fonti
- Qualsiasi attività di ricerca, analisi, monitoraggio che implichi la raccolta di dati quantitativi (e a volte anche qualitativi) richiede un'analisi statistica o se non altro l'utilizzo di statistiche descrittive.
- l'analisi statistica non è utile finché non viene redatta in un formato leggibile e comprensibile.
- un report statistico non è un semplice copia-incolla dei risultati
- In ogni sezione del report deve essere chiara la metodologia utilizzata; le scelte e i risultati devono essere commentati adeguatamente.
- sono consigliati elementi quali
  - ▶ introduzione
  - ▶ review dello stato dell'arte e della letteratura
  - ▶ descrizione delle ipotesi e metodo (comprese fonti dati o metodo di raccolta dei dati)
  - ▶ risultati (metodi e modelli)

# Parte pratica: report statistico e i suoi elementi

- L'introduzione dovrebbe spiegare la motivazione e la natura della ricerca in atto.
- nella review dello stato dell'arte della letteratura si dovrebbe cercare di indicare i principali lavori scientifici o le analoghe ricerche condotte che stanno alla base del proprio lavoro, in modo da giustificare la formulazione delle ipotesi
- Nel metodo va invece descritto come sono stati raccolti i dati, la natura degli **esperimenti** o dei **questionari** o delle **fonti dati**. Se si tratta di dati raccolti direttamente specificare la tipologia di partecipanti e qualsiasi aggiustamento che è stato apportato durante lo svolgimento del progetto.
- La descrizione delle analisi statistiche utilizzate si tende ad includerla nei risultati (è possibile anche nel metodo).

## Parte pratica: report statistico. Altre caratteristiche.

- Nel report statistico è fondamentale indicare quale software è stato utilizzato per l'analisi dei dati, la sua versione e se sono stati utilizzati pacchetti aggiuntivi.
- Inoltre, si dichiara esattamente che tipo di analisi è stata effettuata e le motivazioni sulla scelta dell'uso di alcune analisi e non di altre.
- Tutte queste informazioni sono fondamentali per la riproducibilità della ricerca scientifica e per la valutazione della correttezza dei metodi statistici usati

## Parte pratica: report statistico. Altre caratteristiche.

- Una sezione che dovrebbe essere particolarmente accurata ed estesa, nei report statistici, è la **descrizione dei dati**. Va descritto accuratamente:
  - ▶ campione da cui sono stati presi i dati, utilizzando statistiche descrittive riguardanti ad es. l'età (media, deviazione standard, range), la distribuzione di genere (numero di maschi e femmine), se pertinente, il grado di istruzione (frequenza per titolo di studio, media e/o deviazione standard degli anni di studio), la provenienza geografica.
  - ▶ lo stesso va fatto per i dati più di interesse per gli obiettivi della ricerca

# Un'importante risorsa: il CRAN e un tutorial

Il CRAN (The Comprehensive R Archive Network) disponibile a <https://cran.r-project.org/> offre l'opportunità di vedere i pacchetti disponibili per il software R (che è quello che useremo principalmente in questo corso... dettagli più avanti) organizzati in tematiche o **topics** che vanno dalla statistica bayesiana, passando per l'econometria, il machine learning, finendo in strumenti statistici utili per il Web e l'analisi dei dati testuali o per lo sport analytics.