

# STATISTICA E ANALISI DEI DATI

Introduzione al Corso

---

Dott. Stefano Cirillo  
Dott. Luigi Di Biasi

a.a. 2025-2026

# Docenti del Corso

---



**Stefano Cirillo**

[scirillo@unisa.it](mailto:scirillo@unisa.it)

Ricevimento  
Mercoledì 10:00-11:00  
Laboratorio L8 – Edificio F2,  
Piano -1



**Luigi Di Biasi**

[ldibiasi@unisa.it](mailto:ldibiasi@unisa.it)

Ricevimento  
Giovedì 11:00-12:00  
via Microsoft Teams

# Perché è importante la Statistica e l'Analisi dei Dati?

- La statistica fornisce gli strumenti e metodologie essenziali per **raccogliere, analizzare** e **interpretare** i dati in modo **rigoroso e affidabile:**

La statistica ci indica **come estrarre informazioni** utili da grandi quantità di dati:

Ci aiuta a scegliere campioni rappresentativi;

Correggere eventuali bias;

Minimizzare le fonti di errore;

Generare nuove scoperte attraverso l'analisi dei dati.

La statistica ci mette *al riparo* (nei limiti di un margine di errore considerato accettabile) dal decidere basandoci su *intuizioni* o *ipotesi verificate o supportate dai dati*.

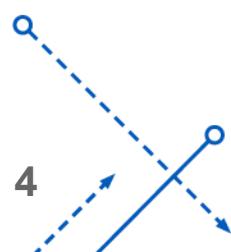
La statistica ci offre gli strumenti per **gestire e quantificare l'incertezza**.

La statistica ci permette di **distinguere i risultati che sono significativi** da quelli che potrebbero essere dovuti al **caso** (argomento del corso saranno i test di ipotesi!).

# SAD $\longleftrightarrow$ AI

---

- La **statistica, l'analisi dei dati e l'intelligenza artificiale** sono strettamente collegate in quanto condividono l'obiettivo di estrarre conoscenza e modelli utili dai dati.
- In linea di principio, **possiamo considerare la statistica come lo «zoccolo duro» dell'AI**.
- La statistica è alla base di diverse tecniche utilizzate nell'intelligenza artificiale (soprattutto nel Machine Learning, una tecnica AI per *imparare automaticamente* dai dati):
  - Tecniche statistiche come la **regressione lineare** e logistica possono essere usati per creare **modelli predittivi** che apprendono dai dati e fanno **previsioni** su nuovi dati.
    - I metodi di apprendimento supervisionato (come **la regressione e la classificazione**) sono legati alla statistica, poiché si cerca di prevedere una variabile target (minimizzando l'errore).
    - L'apprendimento non supervisionato (come il **clustering**) è strettamente connesso all'analisi dei dati, in quanto si cerca di trovare pattern nascosti nei dati senza avere una variabile target.



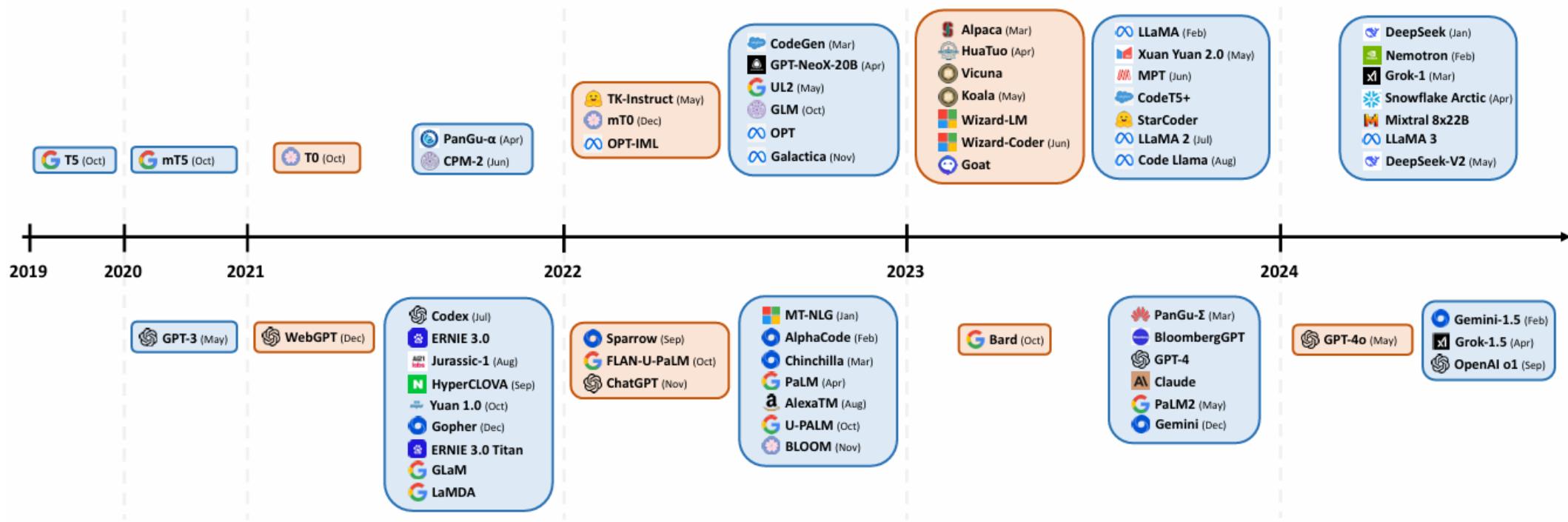
# SAD $\longleftrightarrow$ AI

- **Generazione di dati sintetici:**

- L'AI generativa crea immagini, testi, suoni e dati numerici simulando distribuzioni reali.
- La statistica permette di modellare le distribuzioni e generare campioni realistici.

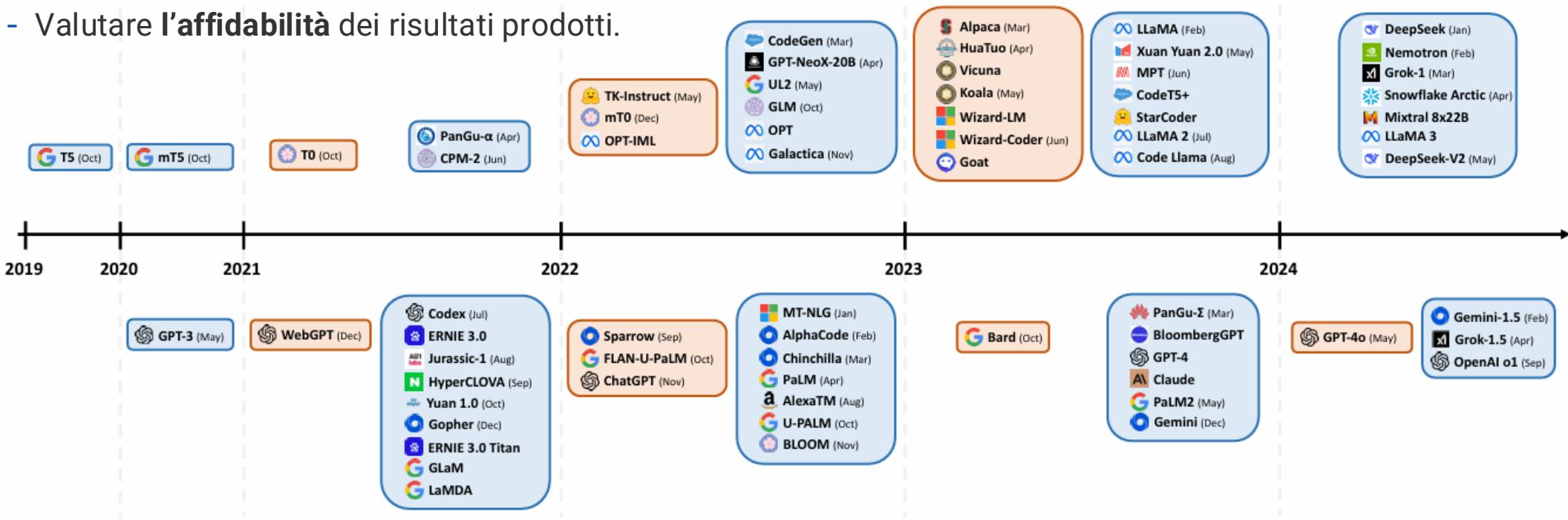
- **Valutazione della qualità dei dati:**

- Serve la statistica per confrontare dati reali e sintetici (es. test di ipotesi, distanze tra distribuzioni)
- Metriche come media, varianza, correlazioni, distribuzioni empiriche sono strumenti chiave.



# SAD $\longleftrightarrow$ AI

- Validazione di **modelli generativi**:
  - Analisi di bias, outlier e varianza nei dati prodotti.
  - Misure di diversità e copertura per evitare **mode collapse** (dati troppo simili).
- Conoscere bene **distribuzioni, correlazioni, test statistici** e tecniche di **visualizzazione** è essenziale per:
  - Capire se un modello generativo **sta funzionando**.
  - Scegliere buoni set di dati **per l'addestramento**.
  - Valutare **l'affidabilità** dei risultati prodotti.



# Obiettivi

---

- Acquisire le **conoscenze teoriche** di metodi e tecniche per il **trattamento** e l'**analisi** di dati complessi, con particolare attenzione alle metodologie di analisi dei dati **più diffuse** e alle **relative applicazioni in contesti informatici**.
- Sviluppare la **capacità di applicare** le conoscenze acquisite a **problemi concreti**.
- Sviluppare la **capacità di comunicare** in modo chiaro, conciso ed efficace **di statistiche dei dati**, anche con interlocutori non esperti di statistica.
- Sviluppare le **competenze pratiche** relative all'analisi statistica dei dati
  - utilizzando uno dei più potenti e flessibili software statistici, ossia il linguaggio di programmazione R (<http://www.R-project.Org/>).
- Sviluppare la capacità di **analizzare in modo critico** problemi complessi, individuando le **variabili chiave** e formulando **ipotesi di soluzione**.



# Prerequisiti

---

- Matematica di base:
  - Nozioni di algebra (equazioni, funzioni, grafici)
  - Concetti fondamentali di probabilità
  - Addizioni, Sottrazioni, ...
- Conoscenza di un linguaggio di programmazione e di un IDE di sviluppo:
  - Solitamente, si richiede familiarità con linguaggi come **R**, Python, o Matlab, utilizzati per analizzare dati
  - Sviluppo del progetto
- Capacità di ragionamento logico:
  - Saper applicare pensiero critico e analitico.
- Conoscenze di base di **calcolo delle probabilità e statistica**.

# Piattaforme di Comunicazione

- Sulla piattaforma di E-Learning è stato creato il canale del corso di Statistica ed Analisi dei Dati
- Sulla piattaforma saranno:
  - Condivise le slide delle lezioni **aggiornate**
  - Inseriti i riferimenti dei Docenti del corso
  - Effettuate comunicazioni relative al corso
  - Inserite le deadline per la sottomissione dei progetti
  - Condivise le dispense di supporto alla comprensione delle tematiche del corso
- Eventuali altre comunicazioni e/o richieste potranno essere fatte tramite email:

[scirillo@unisa.it](mailto:scirillo@unisa.it)

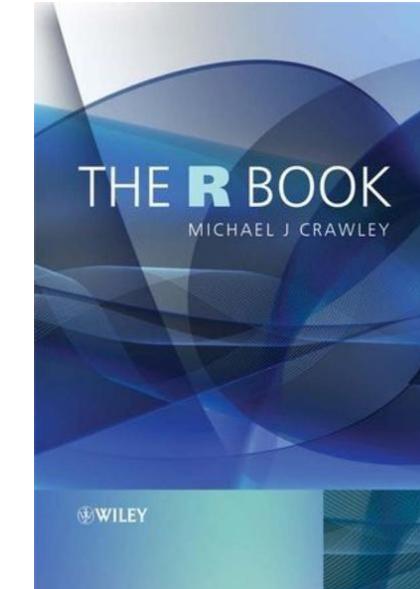
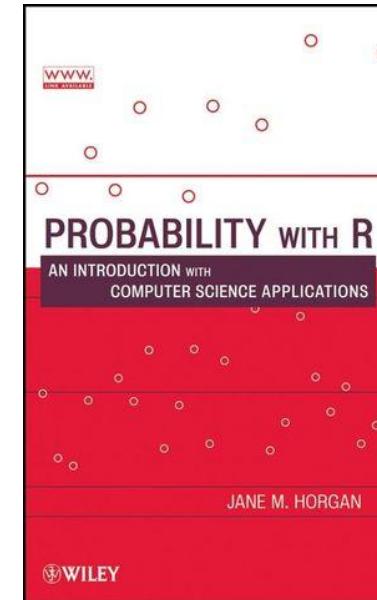
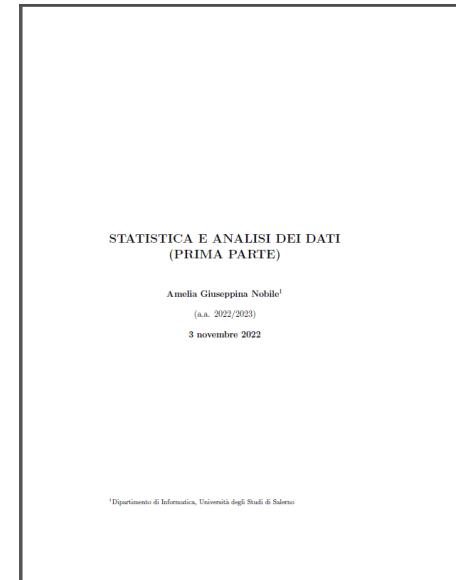
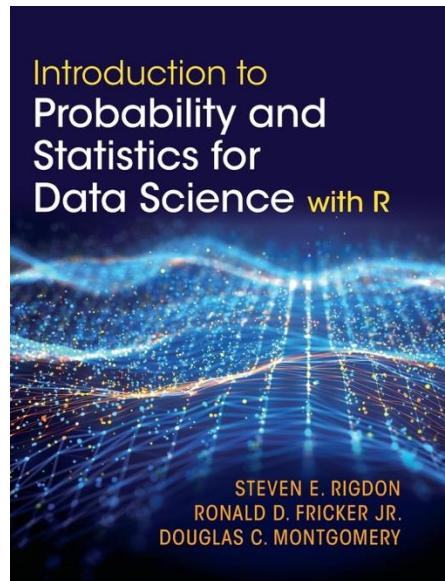
[ldibiasi@unisa.it](mailto:ldibiasi@unisa.it)

The screenshot shows a digital communication platform interface. At the top, a blue header bar displays the text "Primo Semestre - Primo Anno Magistrale 2025/2026 - LM18". Below the header is a large blue diamond-patterned background area. In the bottom right corner of this area is a white button with a black double-headed arrow icon. Below the background, the course title "Statistica ed Analisi dei Dati 2025/2026" is displayed in blue text. On the left side of the main content area, there is a circular profile picture of a man with a beard. To the right of the profile picture, the name "Stefano Cirillo" is written in blue, followed by the word "Docente" in smaller blue text. Below this section, another circular profile picture of a man is shown, with the name "Luigi Di Biasi" to its right and the word "Docente" below it. The overall design is clean and modern, using a color palette of blues, whites, and greys.

# Testi

---

- Steven E. Rigdon, Ronald D. Fricker, Jr, Douglas C. Montgomery (2024) **Introduction to Probability and Statistics for Data Science with R**
- Michael J. Crawley (2012) The R Book, Wiley.
- Jane M. Horgan (2009) Probability With R. An Introduction With Computer Science Applications. Wiley
- Appunti Delle Lezioni Del Docente (Prof.ssa **Amelia Giuseppina Nobile**).



# Argomenti del Corso: Parte 1

---

- L'ambiente integrato R:
  - Principali operatori e funzioni matematiche in R; Vettori; Array e Matrici; Liste; Fattori; Dataframe
- Grafici e tabelle di frequenza:
  - Grafici per vettori numerici; Serie temporali; Distribuzioni di frequenza; Grafici a barre, a bastoncini e diagrammi a torta; Tabelle di contingenza
- Rappresentazioni grafiche dei dati:
  - Istogrammi; Boxplot; Rappresentazioni grafiche per confrontare variabili; Diagramma di Pareto; Grafici di funzioni
- Statistica descrittiva univariata:
  - Funzione di distribuzione empirica; Media, mediana e moda campionarie; Quantili, percentili, decili e quartili; Varianza, deviazione standard e coefficiente di variazione; Momenti campionari e momenti centrati

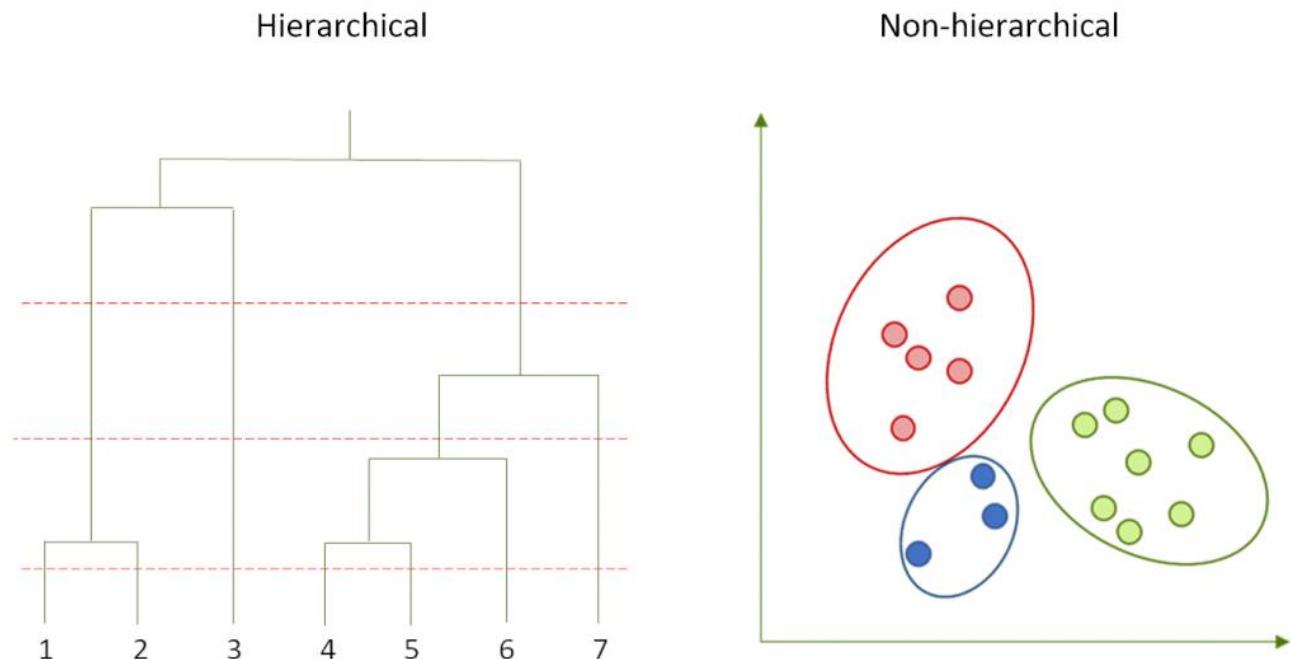
# Argomenti del Corso: Parte 1

- Statistica descrittiva **bivariata**:

- Covarianza e correlazione campionaria; Regressione lineare semplice; Regressione lineare multipla;

- Analisi dei cluster:

- Metrica Euclidea; Altre metriche; Misure di similarità; Misura di non omogeneità totale; Misure di non omogeneità tra cluster; Metodi gerarchici; Analisi del dendrogramma; Metodi non gerarchici;



# Argomenti del Corso: Parte 2

---

- **Variabili Aleatorie Con R:**
  - Distribuzioni Di Probabilità Discrete E Loro Simulazione (Bernoulli, Binomiale, Geometrica E Di Pascal, Ipergeometrica, Di Poisson). Distribuzioni Di Probabilità Continue E Loro Simulazione (Uniforme, Esponenziale, Normale, ...).
- **Statistica Inferenziale Con R:**
  - Stima Puntuale. Proprietà Degli Stimatori. Metodi Per La Ricerca Di Stimatori. Metodi Dei Momenti E Della Massima Verosimiglianza. Stimatori Bayesiani.
- **Stima Per Intervallo Con R:**
  - Intervalli Di Confidenza. Intervalli Di Confidenza Per La Media E Varianza Di Una Popolazione Normale. Intervalli Di Confidenza Di Una Popolazione Di Bernoulli, Di Poisson Ed Esponenziale.
- **Verifica Delle Ipotesi Con R:**
  - Test Sulle Medie. Test Sulle Differenze Tra Medie. Test Sulle Varianze. Test Sulle Proporzioni. Test Chi-quadrato Di Pearson. Test Di Ipotesi In Modelli Di Regressione Lineari E Non Lineari.

# Informazioni per lo sviluppo del progetto

---

- Il Progetto è obbligatorio per tutti
- I gruppi di progetto possono essere formati da massimo **2 Persone**
- Tematiche del progetto:
  - Analisi statistiche su dati reali provenienti da Istituti di Statistica
  - Serie Temporali
  - Analisi su dati provenienti da casi reali di studio
  - Etc.
- I progetti possono essere utilizzati per più corsi accademici
  - Al momento ci stiamo coordinando con alcuni docenti degli altri corsi per permettere di estendere il progetto di questo corso
- Ci sarà parte di una lezione che riguarderà i progetti
  - **Dopo tale lezione, si potranno definire i gruppi ed avviare lo sviluppo dei progetti**

# Modalità d'esame

---

- L'esame consiste di un **progetto** e una **prova orale**
  - Durante la prova orale **verrà presentato il progetto** e verranno **verificate le conoscenze dello studente** rispetto agli argomenti studiati durante il corso
  - Il voto dipenderà dalle conoscenze acquisite e dalla capacità di applicare le metodologie acquisite alla risoluzione di problemi applicativi concreti.
  - Il progetto dovrà essere consegnato almeno 7 giorni prima la data dell'esame sulla piattaforma di e-platform



# Orari del Corso di SAD

[corsi.unisa.it/05225/didattica/orari](http://corsi.unisa.it/05225/didattica/orari)

	Lunedì 22/9	Martedì 23/9	Mercoledì 24/9	Giovedì 25/9
08:30				
09:00	09:00 - 11:00 <b>STATISTICA E ANALISI DEI DATI</b>			
09:30	CIRILLO STEFANO, Di Biasi Luigi Aula F4 [Edificio F2] Lezione			
10:00				
10:30				
11:00				
11:30				
12:00				
12:30				
13:00				
13:30				
14:00				
14:30				
15:00				
15:30			15:30 - 17:30 <b>STATISTICA E ANALISI DEI DATI</b>	15:30 - 17:30 <b>STATISTICA E ANALISI DEI DATI</b>
16:00			CIRILLO STEFANO, Di Biasi Luigi Aula F4 [Edificio F2]	CIRILLO STEFANO, Di Biasi Luigi Aula F4 [Edificio F2]
16:30			Lezione	Lezione
17:00				
17:30				

# DOMANDE?

