

# Elementi di Matematica e di Statistica

## Esercitazione di Statistica

Docente: Riccardo Ievoli  
`riccardo.ievoli@unife.it`

Corso di Laurea in Biotecnologie

Dicembre 2025

# Esercizio 1

## Parte 1

Il dataset “perdite idriche” contiene la percentuale di famiglie che denunciano irregolarità nell'erogazione dell'acqua calcolata per ogni Regione Italiana ( $n = 20$ ) dal 2005 al 2021.

- Si calcolino media e mediana sia per Regione (2005-2021) che per Anno.
- Si calcolino minimo e massimo sia per Regione (2005-2021) che per Anno.
- Si confrontino graficamente gli andamenti temporali di Emilia-Romagna, Campania e Molise.
- Si produca una mappa (di calore) delle perdite idriche nel 2021 e anche un grafico a barre (facoltativo).
- Si calcolino deviazione standard e deviazione standard campionaria sia per Regione che per Anno

# Esercizio 1

## Parte 2

Il dataset “perdite idriche” contiene la percentuale di famiglie che denunciano irregolarità nell'erogazione dell'acqua calcolata per ogni Regione Italiana ( $n = 20$ ) dal 2005 al 2021.

- Si calcolino i coefficienti di variazione sia per Regione che per Anno
- Si effettui un confronto tra gli anni 2015 e 2020 utilizzando il boxplot (facoltativo)
- Si calcoli il coefficiente di correlazione delle perdite idriche di ogni regione rispetto al tempo (variabile indipendente). Quali considerazioni si possono effettuare?
- Si stimino 20 modelli di regressione considerando l'andamento di ogni Regione rispetto al tempo. Quali considerazioni si possono effettuare?

# Esercizio 2

## Parte 1

Il dataset “emissioni” contiene le emissioni di CO2 da trasporto stradale (tonnellate per abitante) dal 1996 al 2003.

Ipotizzando normalità distributiva e ponendo  $\sigma^2 = 0,25$  e quindi noto, si calcoli l'intervallo di confidenza per la media ad un livello di confidenza del 95% per ogni anno.

Ipotizzando normalità distributiva ma considerando  $\sigma^2$  non noto (da stimare per ogni anno), si calcoli l'intervallo di confidenza per la media ad un livello di confidenza del 95% per ogni anno.

# Esercizio 2

## Parte 2

Il dataset “emissioni” contiene le emissioni di CO2 da trasporto stradale (tonnellate per abitante) dal 1996 al 2003.

Ipotizzando normalità distributiva e ponendo  $\sigma^2 = 0,25$  e quindi noto, e dato un livello di significatività  $\alpha = 0,05$  si sottoponga a verifica empirica (per ogni anno di analisi) l'ipotesi nulla tale per cui  $\mu = 2$  contro l'alternativa  $\mu \neq 2$

Ipotizzando normalità distributiva e ponendo  $\sigma^2 = 0,25$  e quindi noto, e dato un livello di significatività  $\alpha = 0,05$  si sottoponga a verifica empirica (per ogni anno di analisi) l'ipotesi nulla tale per cui  $\mu \leq 1,7$  contro l'alternativa  $\mu > 1,7$

Ipotizzando normalità distributiva e ponendo  $\sigma^2 = 0,25$  e quindi noto, e dato un livello di significatività  $\alpha = 0,05$  si sottoponga a verifica empirica (per ogni anno di analisi) l'ipotesi nulla tale per cui  $\mu \geq 1,9$  contro l'alternativa  $\mu < 1,9$

## Esercizio 3

Il numero di goal subiti per partita da una squadra di terza serie, famosa per l'efficacia della sua difesa, segue una distribuzione di Poisson con parametro  $\lambda = 0,25$ .

- Qual è la probabilità che tale squadra non subisca goal nella prossima partita?
- Qual è la probabilità che tale squadra subisca esattamente un goal nella prossima partita?
- Qual è la probabilità che tale squadra subisca meno di due goal nella prossima partita?
- Qual è la probabilità che tale squadra subisca tra uno e due goal nella prossima partita?

## Esercizio 4

Da precedenti rilevazioni è noto che in Italia la probabilità di fallimento delle startup segue una distribuzione binomiale con parametro  $p = 0,2$ .

- Estraendo a caso 5 startup, qual è la probabilità che ne falliscano esattamente 2?
- Estraendo a caso 5 startup, qual è la probabilità che ne fallisca esattamente una?
- Estraendo a caso 5 startup, qual è la probabilità che ne falliscano meno di 3?