

Statistica II (750AA)

Lezione 1

Dario Trevisan – *<https://web.dm.unipi.it/trevisan>*

22/09/2025

Introduzione al corso

Argomenti

0. Richiami di Statistica I (descrittiva, probabilità, test e intervalli)

Argomenti

- 0. Richiami di Statistica I (descrittiva, probabilità, test e intervalli)
- 1. Clustering

Argomenti

- 0. Richiami di Statistica I (descrittiva, probabilità, test e intervalli)
- 1. Clustering
- 2. Classificazione

Argomenti

- 0. Richiami di Statistica I (descrittiva, probabilità, test e intervalli)
- 1. Clustering
- 2. Classificazione
- 3. Regressione

Argomenti

- 0. Richiami di Statistica I (descrittiva, probabilità, test e intervalli)
- 1. Clustering
- 2. Classificazione
- 3. Regressione
- 4. Serie storiche

Argomenti

- 0. Richiami di Statistica I (descrittiva, probabilità, test e intervalli)
 - 1. Clustering
 - 2. Classificazione
 - 3. Regressione
 - 4. Serie storiche
- L'apprendimento di R è centrale nel corso: installate R ed RStudio <https://posit.co/download/rstudio-desktop/>

Materiale didattico

- ▶ MS Team: **750AA 25/26 - STATISTICA II [WIG-LM]**

Troverete:

Materiale didattico

- ▶ MS Team: **750AA 25/26 - STATISTICA II [WIG-LM]**
- ▶ Pagina web del corso
<https://web.dm.unipi.it/trevisan/it/courses/750AA.html>

Troverete:

Materiale didattico

- ▶ MS Team: **750AA 25/26 - STATISTICA II [WIG-LM]**
- ▶ Pagina web del corso
<https://web.dm.unipi.it/trevisan/it/courses/750AA.html>

Troverete:

- ▶ Appunti (slides annotate)

Materiale didattico

- ▶ MS Team: **750AA 25/26 - STATISTICA II [WIG-LM]**
- ▶ Pagina web del corso
<https://web.dm.unipi.it/trevisan/it/courses/750AA.html>

Troverete:

- ▶ Appunti (slides annotate)
- ▶ Demo interattive

Materiale didattico

- ▶ MS Team: **750AA 25/26 - STATISTICA II [WIG-LM]**
- ▶ Pagina web del corso
<https://web.dm.unipi.it/trevisan/it/courses/750AA.html>

Troverete:

- ▶ Appunti (slides annotate)
- ▶ Demo interattive
- ▶ R notebooks (codice e commenti)

Materiale didattico

- ▶ MS Team: **750AA 25/26 - STATISTICA II [WIG-LM]**
- ▶ Pagina web del corso
<https://web.dm.unipi.it/trevisan/it/courses/750AA.html>

Troverete:

- ▶ Appunti (slides annotate)
- ▶ Demo interattive
- ▶ R notebooks (codice e commenti)
- ▶ RegISTRAZIONI delle lezioni

Materiale didattico

- ▶ MS Team: **750AA 25/26 - STATISTICA II [WIG-LM]**
- ▶ Pagina web del corso
<https://web.dm.unipi.it/trevisan/it/courses/750AA.html>

Troverete:

- ▶ Appunti (slides annotate)
- ▶ Demo interattive
- ▶ R notebooks (codice e commenti)
- ▶ Registrazioni delle lezioni
- ▶ Test d'esame

Orario lezioni

- ▶ Lunedì F4 14:30 - 16:30

Orario lezioni

- ▶ Lunedì F4 14:30 - 16:30
- ▶ Mercoledì B25 8:30 - 11:30 (portate computer/tablet)

Ricevimento

► Giorno:

Ricevimento

- ▶ Giorno:
- ▶ Orario:

Ricevimento

- ▶ Giorno:
- ▶ Orario:
- ▶ Comunque su appuntamento, anche su Teams

Ricevimento

- ▶ Giorno:
- ▶ Orario:
- ▶ Comunque su appuntamento, anche su Teams
- ▶ Contattatemi **sempre** via mail o messaggio su Teams.

Modalità di esame

- ▶ **prova teorica:** test scritto con domande sugli argomenti del corso (vero/falso da argomentare brevemente) durata 1 ora.

Modalità di esame

- ▶ **prova teorica:** test scritto con domande sugli argomenti del corso (vero/falso da argomentare brevemente) durata 1 ora.
- ▶ **prova pratica:** progetti autonomi di analisi statistica di dati reali, su **almeno due dei tre argomenti principali** (classificazione, regressione, serie storiche). Vedere foglio di indicazioni generali sulle relazioni.

Modalità di esame

- ▶ **prova teorica:** test scritto con domande sugli argomenti del corso (vero/falso da argomentare brevemente) durata 1 ora.
- ▶ **prova pratica:** progetti autonomi di analisi statistica di dati reali, su **almeno due dei tre argomenti principali** (classificazione, regressione, serie storiche). Vedere foglio di indicazioni generali sulle relazioni.
- ▶ Il voto finale tiene conto di entrambe le prove.

Statistica descrittiva

Indicatori di centralità e variabilità

Date osservazioni (un campione statistico) $\{x_i\}_{i=1}^n$ definiamo gli indicatori di centralità:

- ▶ moda

e gli indicatori di variabilità (o dispersione):

Indicatori di centralità e variabilità

Date osservazioni (un campione statistico) $\{x_i\}_{i=1}^n$ definiamo gli indicatori di centralità:

- ▶ moda
- ▶ mediana

e gli indicatori di variabilità (o dispersione):

Indicatori di centralità e variabilità

Dato osservazioni (un campione statistico) $\{x_i\}_{i=1}^n$ definiamo gli indicatori di centralità:

- ▶ moda
- ▶ mediana
- ▶ media

e gli indicatori di variabilità (o dispersione):

Indicatori di centralità e variabilità

Date osservazioni (un campione statistico) $\{x_i\}_{i=1}^n$ definiamo gli indicatori di centralità:

- ▶ moda
- ▶ mediana
- ▶ media

e gli indicatori di variabilità (o dispersione):

- ▶ quantili (quartili, percentili)

Indicatori di centralità e variabilità

Date osservazioni (un campione statistico) $\{x_i\}_{i=1}^n$ definiamo gli indicatori di centralità:

- ▶ moda
- ▶ mediana
- ▶ media

e gli indicatori di variabilità (o dispersione):

- ▶ quantili (quartili, percentili)
- ▶ varianza (deviazione standard)

Moda

Mediana

Media

Varianza

Funzioni di Perdita/Costo/Errore

La media $\bar{x}_n$ minimizza il rischio quadratico (MSE)

$$\bar{x}_n \in \arg \min_{h \in \mathbb{R}} \sum_{i=1}^n (x_i - h)^2$$

- Introduciamo nuove funzioni di perdita/costo/errore $\ell(x; h)$ e **minimizziamo il rischio empirico** (ERM):

$$\arg \min_{h \in \mathbb{R}} \sum_{i=1}^n \ell(x_i; h).$$

Errore assoluto

Errore assoluto asimmetrico

Errore 0-1

Errore esponenziale assoluto

WERMS



► Giocate voi! <https://dario-trevisan.shinyapps.io/WERMS/>

WERMS



- ▶ Giocate voi! <https://dario-trevisan.shinyapps.io/WERMS/>
- ▶ Codice (R shiny) disponibile nel team e nella pagina del corso

Dati multivariati

Distanze tra punti nello spazio

Nella statistica multivariata si osserva

- ▶ un campione di n individui (taglia): x_i , per $i = 1 \dots, n$

Possiamo organizzare i dati in una tabella (*data frame* in R).

Distanze tra punti nello spazio

Nella statistica multivariata si osserva

- ▶ un campione di n individui (taglia): x_i , per $i = 1 \dots, n$
- ▶ e d caratteristiche (features) per ciascun individuo $x_{i,j}$, per $1, \dots, d$.

Possiamo organizzare i dati in una tabella (*data frame* in R).

Distanze tra punti nello spazio

Confrontare punti nello spazio $\mathbb{R}^d$ è cruciale per l'analisi di dati multivariati. Distanze comuni:

- ▶ Euclidea:

Distanze tra punti nello spazio

Confrontare punti nello spazio $\mathbb{R}^d$ è cruciale per l'analisi di dati multivariati. Distanze comuni:

- ▶ Euclidea:
- ▶ Manhattan (ℓ^1 , taxicab):

Distanze tra punti nello spazio

Confrontare punti nello spazio $\mathbb{R}^d$ è cruciale per l'analisi di dati multivariati. Distanze comuni:

- ▶ Euclidea:
- ▶ Manhattan (ℓ^1 , taxicab):
- ▶ ℓ^p (Minkowski):

Distanze tra punti nello spazio

Confrontare punti nello spazio $\mathbb{R}^d$ è cruciale per l'analisi di dati multivariati. Distanze comuni:

- ▶ Euclidea:
- ▶ Manhattan (ℓ^1 , taxicab):
- ▶ ℓ^p (Minkowski):
- ▶ ℓ^∞ (del massimo):

Centroide e Medoide

Medoids!



- ▶ Giocate voi! <https://dario-trevisan.shinyapps.io/medoids/>

Medoids!



- ▶ Giocate voi! <https://dario-trevisan.shinyapps.io/medoids/>
- ▶ Codice (R shiny) disponibile nel team e nella pagina del corso