



Università degli Studi di Salerno

Dipartimento di Ingegneria dell'Informazione ed Elettrica e
Matematica Applicata (DIEM)

Relazione di progetto

Sviluppo di un modello di regressione lineare su dataset

Corso di Statistica Applicata - A.A. 2024/25

Studenti Gruppo 16:

Corradomaria Giachetta

Matricola: 0612708054

Francesco Peluso

Matricola: 0612707469

Gerardo Selce

Matricola: 0612707692

Anuar Zouhri

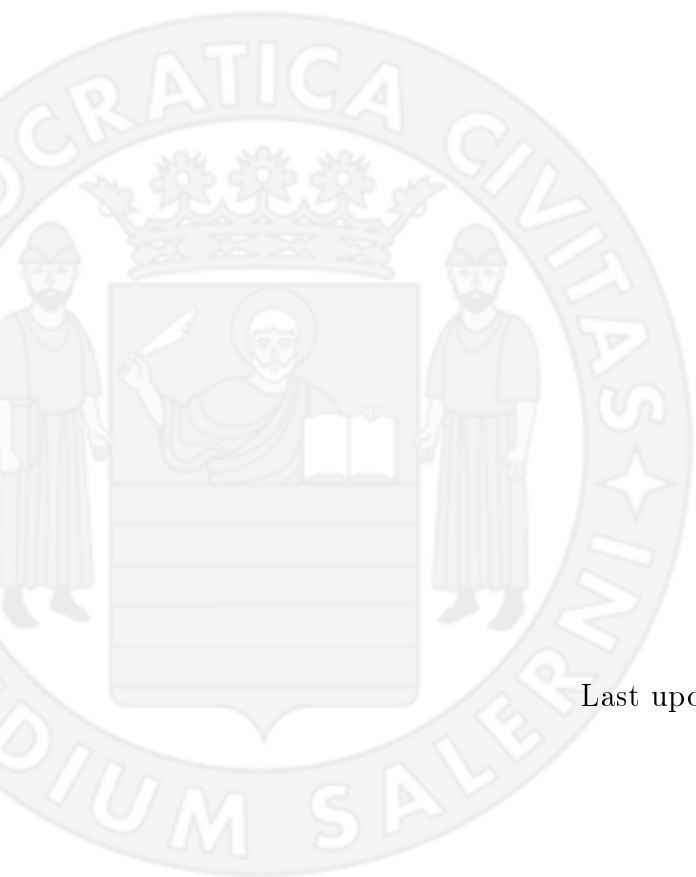
Matricola: 0612707505

Docenti:

Prof. Fabio Postiglione

Prof. Paolo Adesso

Last update: 17 giugno 2025



Indice

1	Descrizione del dataset fornito	2
2	Analisi delle caratteristiche del dataset	3
2.1	Boxplot dei dati	3
2.2	Analisi di normalità	4



1 Descrizione del dataset fornito

A completezza del progetto si riporta la descrizione del dataset da analizzare.

Variabile dipendente

y_VideoQuality → Qualità percepita del video

Tale indice è immaginato come frutto di una opportuna trasformazione di un punteggio assegnato a un campione di immagini da volontari che compilano un questionario. Esso sarà funzione di diverse caratteristiche proprie dei video, tra cui:

- la presenza o meno di rumore;
- la presenza o meno di *motion blur*;
- la nitidezza;
- la profondità di campo;
- la risoluzione;
- le aberrazioni ottiche visibili;
- la gamma dinamica;
- la fedeltà cromatica.

Variabili indipendenti (regressori)

Sono delle quantità di cui l'operatore ha il controllo (parziale o totale) selezionando:

- l'attrezzatura video da utilizzare;
- i parametri di ripresa.

Rappresentano indici standardizzati:

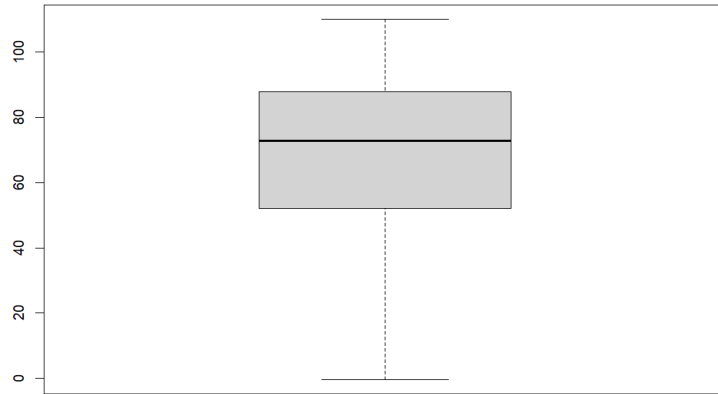
- **x1_ISO** → ISO (sensibilità del sensore)
- **x2_FRatio** → Rapporto Focale
- **x3_Time** → Tempo di Esposizione (in relazione al frame rate utilizzato)
- **x4_MP** → Megapixel del sensore
- **x5_CROP** → Fattore di Crop
- **x6_FOCAL** → Focale
- **x7_PixDensity** → Densità di pixel

2 Analisi delle caratteristiche del dataset

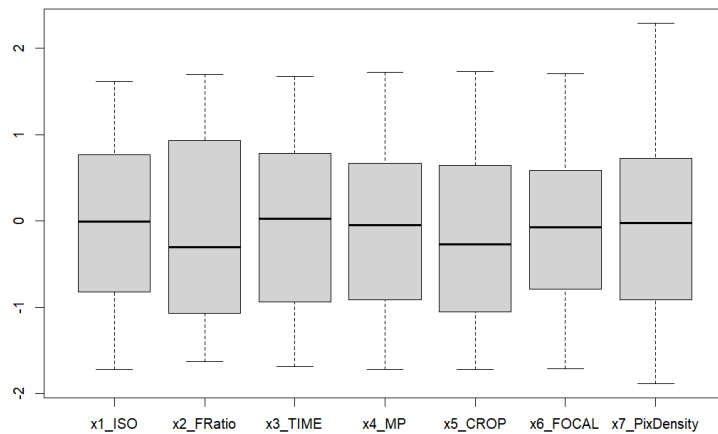
In questa fase preliminare si illustreranno le principali considerazioni fatte sul dataset fornito.

2.1 Boxplot dei dati

Si considerino i seguenti boxplot delle variabili del dataset.



(a) Boxplot della variabile dipendente $y_VideoQuality$



(b) Boxplot delle variabili indipendenti x_i

Figura 1: Boxplot delle variabili considerate

Si osservi innanzitutto che i valori per ciascuna variabile sono tutti contenuti all'interno dell'intervallo interquartile e che quindi non sono presenti outliers. Per quel che riguarda la variabile dipendente $y_VideoQuality$ si è osservato che il valore della media e della mediana sono simili, infatti valgono rispettivamente $media = 72.8135$, $mediana = 68.6081$. Si è osservato inoltre che i valori assunti dalla variabile $x7_PixDensity$ coprono un intervallo maggiore rispetto alle altre variabili indipendenti.

2.2 Analisi di normalità

Anche se non strettamente necessario ai fini del metodo di regressione, si è comunque deciso di verificare se qualcuna delle variabili indipendenti avesse una distribuzione normale. Tra i diversi qq-plot, si osserva che la variabile `x6_Focal` sembrerebbe avere una

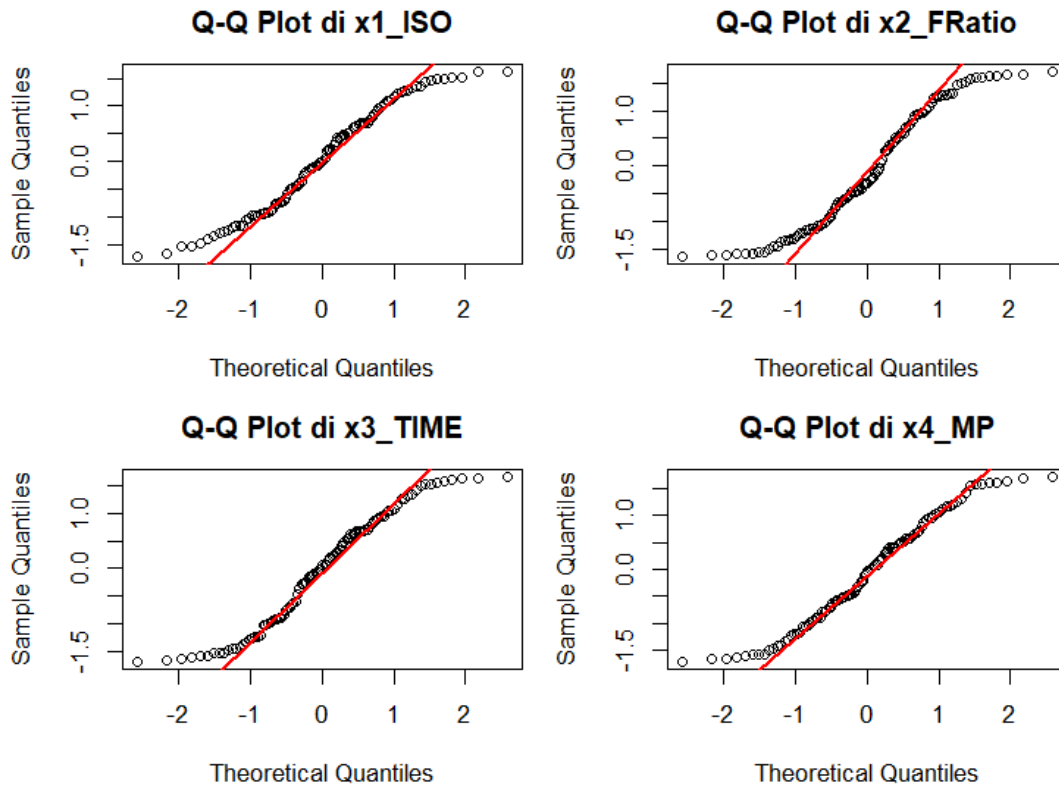
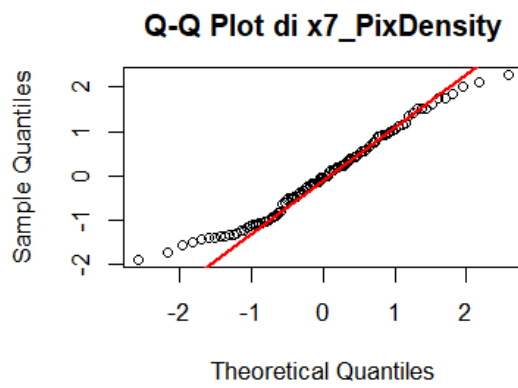
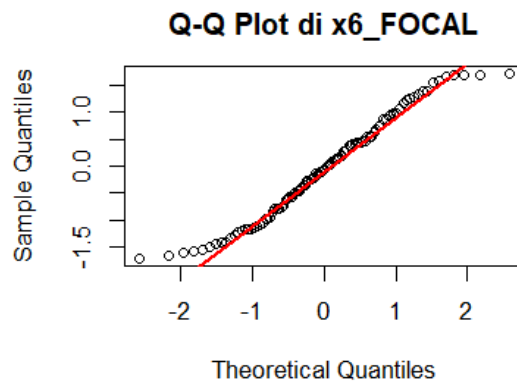
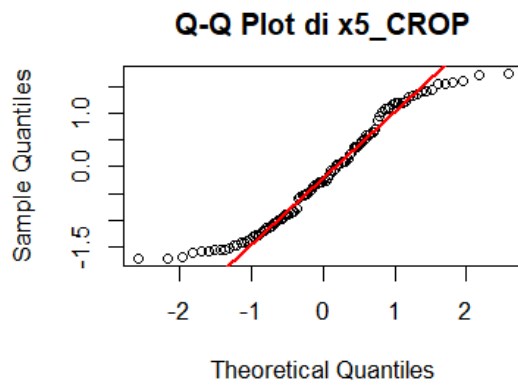


Figura 2:

distribuzione normale. Applicando il test di shapiro a questa variabile si ottiene

$$W = 0.97, p\text{-value} = 0.02.$$

Il valore di p-value ottenuto non si discosta molto da 0.05 e si potrebbe perciò supporre che la variabile sia distribuita come una normale.



3 Analisi della dipendenza tra le variabili

