



SCHOOL OF ADVANCED SCIENCES

Winter Semester 2024-2025

Digital Assessment 3 : Residual Analysis

Programme Name & Branch : M.Sc. & Data Science
Slot : L29+L30
Course Name & code : Regression Analysis and Predictive Models Lab & PMDS504P
Faculty Name : Dr. Jisha Francis
Due Date & Max. Marks : 15 March, 2025 & 10 Marks

Instructions:

1. Answer all the questions provided.
2. Upload a well-structured PDF with your analysis, observations, conclusions, and relevant outputs.

Problem Statement

You are given a dataset containing various health-related variables for 20 individuals. Your task is to analyze the relationship between **Diastolic Blood Pressure (BP)** and other predictor variables using **simple and multiple linear regression techniques**.

Dataset

The dataset consists of the following variables:

- **Pt**: Patient ID
- **BP**: Diastolic Blood Pressure
- **Age**: Age of the individual
- **Weight**: Body weight (kg)
- **BSA**: Body Surface Area (m²)
- **Duration**: Duration of illness (years)
- **Pulse**: Pulse rate (beats per minute)
- **Stress**: Stress level (arbitrary scale)

Pt	BP	Age	Weight	BSA	Duration	Pulse	Stress
1	105	47	85.4	1.75	5.1	63	33
2	115	49	94.2	2.10	3.8	70	14
3	116	49	95.3	1.98	8.2	72	10
4	117	50	94.7	2.01	5.8	73	99
5	112	51	89.4	1.89	7.0	72	95
6	121	48	99.5	2.25	9.3	71	10
7	121	49	99.8	2.25	2.5	69	42
8	110	47	90.9	1.90	6.2	66	8
9	110	49	89.2	1.83	7.1	69	62
10	114	48	92.7	2.07	5.6	64	35
11	114	47	94.4	2.07	5.3	74	90
12	115	49	94.1	1.98	5.6	71	21
13	114	50	91.6	2.05	10.2	68	47
14	106	45	87.1	1.92	5.6	67	80
15	125	52	101.3	2.19	10.0	76	98

16	114	46	94.5	1.98	7.4	69	95
17	106	46	87.0	1.87	3.6	62	18
18	113	46	94.5	1.90	4.3	70	12
19	110	48	90.5	1.88	9.0	71	99
20	122	56	95.7	2.09	7.0	75	99

Instructions

1. Data Preparation

- Load the dataset from `bloodpress.csv`.
- Extract the following variables:
 - **Response Variable (Y):** BP (Diastolic Blood Pressure)
 - **Predictors (X):** Age, Weight, and Duration

2. Simple Linear Regression

- Perform separate linear regressions to model BP as a function of each predictor (Age, Weight, and Duration).
- Display the regression summary and scatter plots for each case.
- Plot the residuals against Weight for the model $\mathbf{BP} \sim \mathbf{Age}$.

3. Multiple Linear Regression

- Perform a multiple linear regression to model BP as a function of **Age and Weight**.
- Display the regression summary.
- Plot the residuals of this model against **Duration**.

4. Interpretation & Observations

- Analyze the regression summaries and discuss:
 - The significance of predictor variables.
 - Goodness-of-fit (R^2 value).
 - Residual behavior and potential model improvements.