

STATISTICS AND DATA ANALYSIS

PREPARED BY

ASSOCIATE PROF. DR NURULHUDA FIRDAUS BT MOHD AZMI

INTELLIGENT INFORMATICS

FACULTY OF ARTIFICIAL INTELLIGENCE

UNIVERSITI TEKNOLOGI MALAYSIA

LEARNING OUTCOME

At the end of the course, participants will be able to:

- **Understand the fundamental principles of statistics and data analytics.**
- **Apply basic statistical techniques to analyse and interpret data effectively.**
- **Utilise simple data analytics tools to explore and visualize data for decision-making purposes.**

COURSE OUTLINE



BACIS STATISTICS AND DATA ANALYTICS UNDERSTANDING



DESCRIPTIVE: EXPLORING AND SUMMARIZING DATA
SPSS FOR DESCRIPTIVE STATISTICS
PROBABILITY BASIC

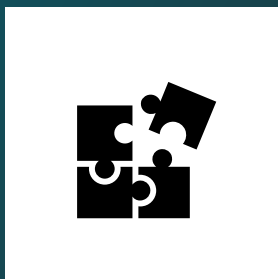


INFERENTIAL: MAKING DECISION FROM SAMPLE TO POPULATION
SPSS FOR STATISTICAL TEST (PARAMETRIC AND NON-PARAMETRIC)



RELATIONSHIP & PREDICTION
SPSS FOR CORRELATION (PARAMETRIC AND NON-PARAMETRIC)
BASIC DATA MINING
SPSS FOR REGRESSION

FIRST SKILL: SPSS for Descriptive Statistics



Visual Way for Descriptive Statistics

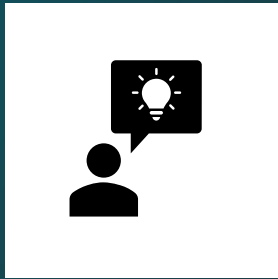


Numerical Way: Central Tendency & Data Position



Numerical Way: Data Variation & Shape of Distribution

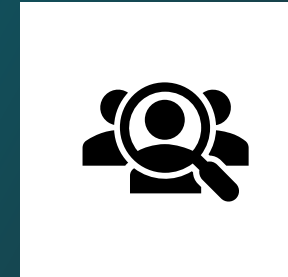
SECOND SKILL: SPSS for Statistical Test



One Sample Statistical
Test: Parametric and Non-
Parametric



Independent Group
Statistical Test: Parametric
and Non-Parametric

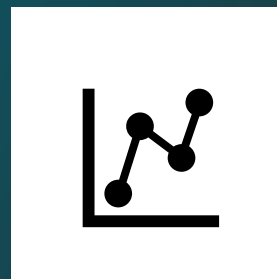


Dependent (Paired) Group
Statistical Test: Parametric
and Non-Parametric

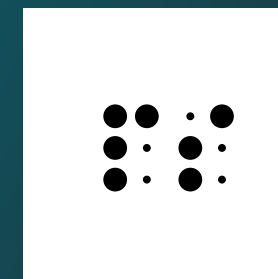
THIRD SKILL: SPSS for Correlation



Pearson Correlation:
Parametric

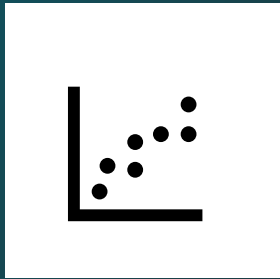


Spearman Correlation:
Non- Parametric

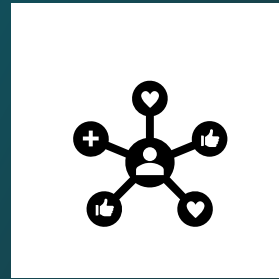


Cramer's V and Phi
Correlation: Qualitative
Data

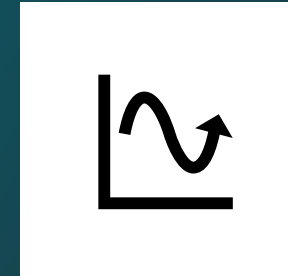
FOURTH SKILL: SPSS for Regression



Simple Linear Regression



Multiple Regression



Polynomial Regression



ACADEMIC BACKGROUND:

Diploma in Computer Science (UTM) (1997), BSc Computer Science (UTM) (1999),

MSc Applied Statistics (UPM) (2001) - Thesis Title: Identifying Multiple Outliers In Linear Regression By Clustering Methodology Using MM Estimators

PhD in Computer Science (Univ. Of York, UK)(2014) - Thesis Title: Artificial Immune Systems for Information Filtering: Focusing on Profile Adaptation

AREA OF RESEARCH:

Soft Computing, Operational Research, Data Mining, Information Filtering, Statistical Data Analysis

Research Group (RG): Operational Business Intelligence (OBI)

Interest Group (IG): Machine Learning for Data Science (MLDS)

CONTACT INFORMATION:

03-22031390, +60125719624 Email: huda@utm.my

Course SCHEDULE

MEETING 1

Ice Breaking
Introduction
Descriptive Statistics
Probability Basic
Active Learning Activities

MEETING 2

Normality Test
Parametric Statistical Test
Active Learning Activities

MEETING 3

Non-Parametric Statistical Test

Correlation Parametric
Active Learning Activities

MEETING 4

Correlation Non-Parametric

Teaser to Data Mining

Active Learning Activities

MEETING 5

Regression Analysis for Prediction (Simple, Multiple & Polynomial)
Active Learning Activities

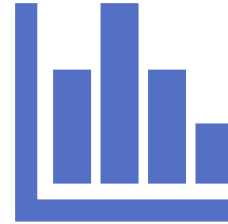
EVERY 1 HOUR CLASS WILL HAVE 10 MINS BREAK

CLASS BEGIN: 9.30 AM

LUNCH BREAK: 1.00 PM – 2.30 PM

CLASS DISMISS: 4.00 PM

SOFTWARE WHICH WILL BE USE



IBM SPSS (STATISTICAL PACKAGE FOR SOCIAL SCIENCE)

Descriptive Statistics, Normality Test, Statistical Test, Correlation Analysis, Regression Analysis

MICROSOFT EXCEL

Descriptive Statistics, Correlation Analysis, Regression Analysis

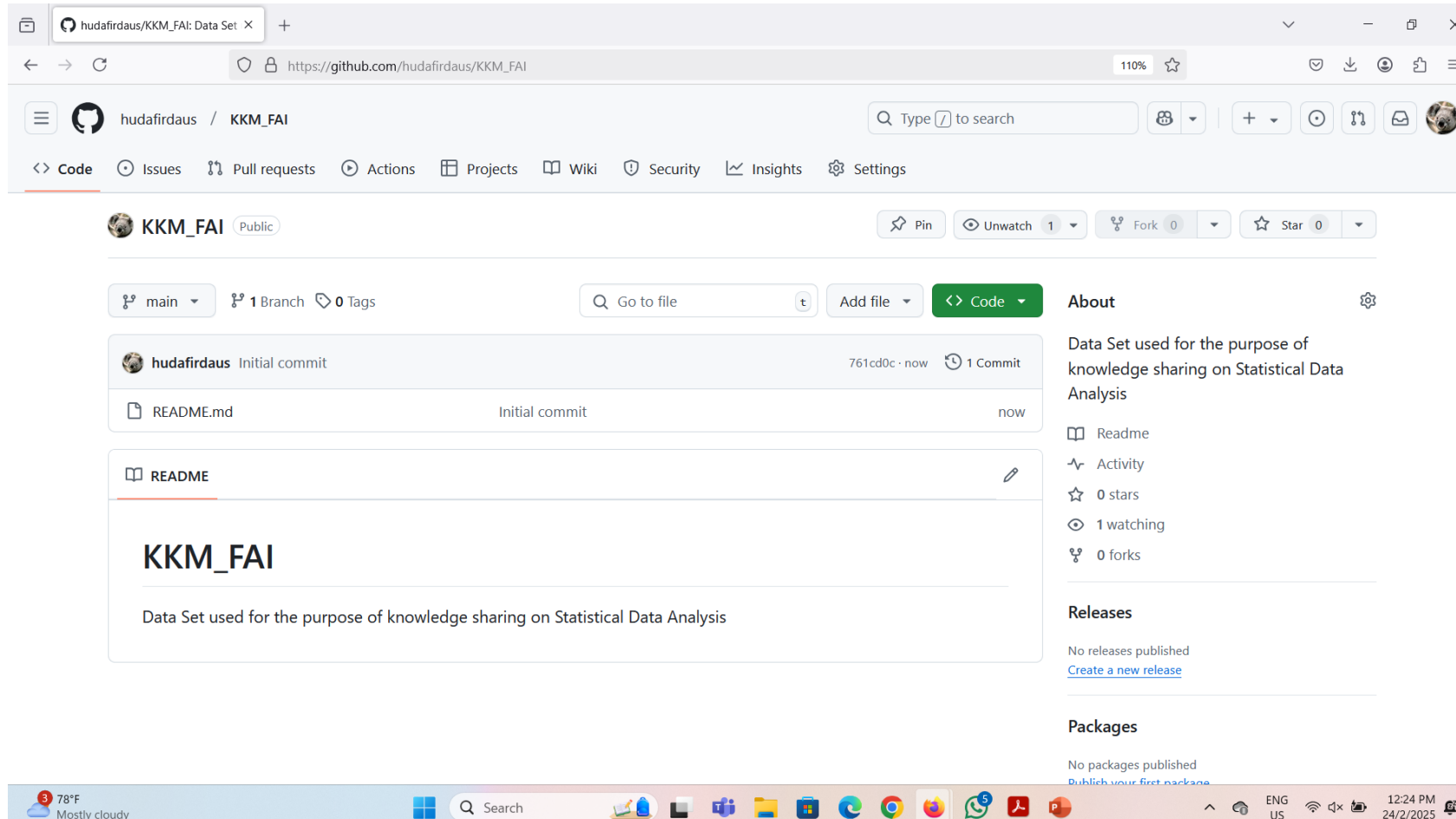
Need to add-ins Analysis-Tool Pack (Excel -> Option -> Add Ins)

OTHERS: YOU CAN USE ANY SOFTWARE YOU PREFER TO CONDUCT STATISTICAL ANALYSIS (FOR INSTANCE, PYTHON, R-PROGRAMMING, MATLAB, STRATA AND ETC)

DATA SET USED IN CLASS

LINK FOR DATA SET:

https://github.com/hudafirdaus/KKM_FAI



The screenshot shows a web browser displaying the GitHub repository page for 'KKM_FAI' by user 'hudafirdaus'. The repository is public and has 1 branch (main) and 0 tags. The commit history shows an initial commit by 'hudafirdaus' with the message 'Initial commit' and a file named 'README.md'. The README content is titled 'KKM_FAI' and describes the data set as 'Data Set used for the purpose of knowledge sharing on Statistical Data Analysis'. The repository has 0 stars, 1 watching, and 0 forks. The 'About' section states: 'Data Set used for the purpose of knowledge sharing on Statistical Data Analysis'. The 'Releases' section indicates 'No releases published' with a link to 'Create a new release'. The 'Packages' section indicates 'No packages published' with a link to 'Publish your first package'.



*“Life is a journey
with problems to
solve and lessons to
learn but most of
all experience to
enjoy”*

