RED & WHITE SINCE2008
SKILL EDUCATION
SHAPING SKILLS FOR SCALING HIGHER...!!!

Title: Derivable Judgement: A Statistical Decision-Making Model
Duration: 6 Hours
Type: Theory + Practical

🎯 Objective:

To help students apply inferential statistics to evaluate multiple hypotheses based on a given dataset. This will test their theoretical understanding and practical application of statistical concepts including hypothesis testing, confidence intervals, statistical errors, and various tests (z, t, chi-square, ANOVA), as well as correlation and covariance.

Project Scenario:

You are a data analyst at a public health research organization. You've been given a dataset that contains health records of individuals from different regions, categorized by gender, age groups, lifestyle habits (smoking, exercise), and diagnosed diseases (e.g., Diabetes, Hypertension). Your task is to use statistical techniques to **derive judgements** about significant factors affecting disease occurrence.

🧩 Dataset Structure (Generate from an AI Tool)

Field Name	Data Type	Description
record_id	UUID/String	Unique identifier for each health record
age_group	String	Age group category (e.g., "18-25", "26-35", "36-45", "46-60", "60+")
age	Int	Age of individuals
weight	Int	Weight of Individual

gender	String	Gender of the individual (e.g., "Male", "Female", "Other")
region	String	Geographic region (e.g., "North", "South", "East", "West")
smoking_status	String	Smoking habit (e.g., "Smoker", "Non-Smoker", "Former Smoker")
exercise_frequency	String	Frequency of physical exercise (e.g., "Daily", "Weekly", "Rarely", "Never")
bmi	Float	Body Mass Index of the individual
blood_pressure	Float	Measured systolic blood pressure (mmHg)
diabetes	Boolean	Whether the individual has diabetes (True/False)
hypertension	Boolean	Whether the individual has hypertension (True/False)
cholesterol_level	Float	Total cholesterol level (mg/dL)
glucose_level	Float	Fasting glucose level (mg/dL)
visit_date	Date	Date of health check-up or diagnosis

📌 Task Checklist

- **Part A – Theoretical Foundation (Short Notes & Explanation):**

Write a short conceptual note (2–3 lines each) covering

1. What is inferential statistics?
2. What is hypothesis testing and its components?
3. Explain confidence interval and critical value.
4. Define p-value.
5. Differentiate Type I and Type II errors.
6. Brief descriptions of z-test, t-test, chi-square test, and ANOVA test.
7. What is Covariance?
8. What is Correlation?

- **Part B – Data Analysis & Testing Tasks:**

Perform the following using Python (preferably with NumPy, Pandas, SciPy, or Statsmodels):

1. **Formulate at least two hypotheses** from the dataset. Example:
    - H_0: *Smoking has no effect on Diabetes prevalence.*
    - H_1: *Smoking affects Diabetes prevalence.*
2. **Calculate Confidence Intervals** for key numerical data like age, weight, etc.
3. **Find the Critical Value** and **p-value** to interpret the test results.
4. **Perform z-test or t-test** based on sample size (mean comparison across groups).
5. **Conduct a chi-square test** on categorical data (e.g., Smoking habit vs Disease).
6. **Perform an ANOVA test** to check if age groups significantly differ in disease rate.
7. **Calculate Covariance and Correlation** between continuous variables (e.g., Age vs. BMI).
8. Clearly **state the result and interpretation** (Accept/Reject H_0) of each test performed.

✅ Submission Guidelines

- Include practical implementation in a Jupyter Notebook, Excel/Sheets, or screenshots.
- Label all charts clearly and write short interpretations under each result.
- **GitHub Repository:**
    - Create a GitHub repository to host your project.
    - Upload your project files, including source code, and documentation to the repository.
    - Add a document (PDF) explaining theory concepts with definitions.
    - Ensure that you provide a clear and descriptive README.md file.

Remember to follow the instructions provided professionally, make suitable assumptions wherever necessary, and avoid copying code or content from unauthorized sources. Good luck with your project work!

Derivable Judgement

Mathematics & Advanced Statistics

BRING ON YOUR CODING ATTITUDE