

Assignment-2: Hypothesis Testing

Objective

Perform 5 statistical hypothesis tests to explore if certain sensor measurements or their statistical aggregates differ significantly between different operational phases or correlate with Remaining Useful Life (RUL) in the preprocessed data.

Dataset Links

- [Metadata with RUL](#)
- [Rolling Window Sequence](#)

Assignment Tasks

1. Data Preparation

Use the CMAPSS dataset after preprocessing, which includes sensor metadata, computed RUL values, and rolling window sequences representing sensor readings over cycles.

2. Formulate Hypotheses

Example:

- Null hypothesis (H0): The mean value of sensor 7 during the first 30% of operational cycles is equal to the mean value during the last 30% of cycles.
- Alternative hypothesis (H1): The mean value of sensor 7 differs significantly between the first and last 30% of cycles.

3. Select an Appropriate Test

- Use a two-sample t-test if sensor data is normally distributed.
- Otherwise, use a Mann-Whitney U test for non-parametric data.

4. Execute the Test

Calculate the test statistic and p-value comparing sensor 7 readings in early vs. late operational cycles using rolling window sequence data.

5. Interpret Results

If $p\text{-value} < 0.05$, reject H0, indicating a statistically significant difference in sensor behavior, suggesting sensor 7's sensitivity to engine degradation phases.

6. Extended Analysis

Test correlation between aggregated sensor features (mean, standard deviation) and RUL using Pearson or Spearman correlation tests.

7. Report

Document hypotheses, test choices, results, and discuss implications for predictive maintenance.

Deliverables

- A **Jupyter Notebook** implementing all 5 hypothesis tests with:
 - Clean, well-documented code for data loading, preprocessing, hypothesis testing, and correlation analysis.
 - Statistical test results displayed with p-values, test statistics, and interpretation notes.
 - Visualizations (e.g., boxplots, histograms, scatter plots) aiding exploratory data analysis and result explanation.
 - Final summary report within the notebook, compiling all findings and their significance regarding RUL prediction.
- The notebook should be saved with all outputs included (`.ipynb` file with output cells saved) for easy review and reproducibility.

This assignment equips learners with concrete skills in applying hypothesis testing on time series and metadata from predictive maintenance datasets, strengthening critical thinking for feature evaluation and monitoring.