# NORMALITY TEST

## INFERENTIAL STATISTICS

*prepared by:*

**Gyro A. Madrona**
Electronics Engineer

# TOPIC OUTLINE

**Shapiro-Wilk**

**Anderson-Darling**

# SHAPIRO-WILK

# SHAPIRO-WILK

The Shapiro-Wilk normality test calculates a statistic based on the correlation between the data and the corresponding normal distribution; effective for sample size ($\boldsymbol{n \leq 50}$).

**Shapiro Function**

`shapiro()` is a statistical tool used to perform the Shapiro-Wilk test for normality.

Syntax

```
w_stat, p_value = stats.shapiro(data)
```

Null Hypothesis

$\boldsymbol{H_o}$: **p-value** $\geq \boldsymbol{\alpha}$ (Normal data)

Alternative Hypothesis

$\boldsymbol{H_a}$: **p-value** $< \boldsymbol{\alpha}$ (Non-normal data)

# EXERCISE

Perform **Shapiro-Wilk** normality test for the given dataset.

Dataset

defects-data-30-samples.csv

Solution

```
w_stat, p_value =
stats.shapiro(df['Defects'])
```

Null Hypothesis

$H_o$: **p-value** $\geq 0.05$ (Normal data)

Alternative Hypothesis

$H_a$: **p-value** $< 0.05$ (Non-normal data)

# ANDERSON-DARLING

# ANDERSON-DARLING

The Anderson-Darling normality test provides a more sensitive test by giving weight to the tails of the distribution; effective for larger sample size ($\boldsymbol{n > 50}$).

**Anderson Function**

`anderson()` is a statistical tool used to perform the Aderson-Darling test for normality.

Syntax

```
a2_stat, critical_values, alpha =
stats.aderson(data)
```

Null Hypothesis

$\boldsymbol{H_o}$: $\boldsymbol{A^2 \leq}$ **critical value** (Normal data)

Alternative Hypothesis

$\boldsymbol{H_a}$: $\boldsymbol{A^2 >}$ **critical value** (Non-normal data)

# EXERCISE

Perform **Anderson-Darling** normality test for the given dataset.

Dataset

defects-dataset.csv

Solution

```
a2_stat, critical_values, alpha =
stats.anderson(df['Defects'])
```

at $\alpha = 0.05$

Null Hypothesis

$H_o$: $A^2 \leq$ **critical value** (Normal data)

Alternative Hypothesis

$H_a$: $A^2 >$ **critical value** (Non-normal data)

# LABORATORY