NORMALITY TEST

HYPOTHESIS TESTING

prepared by:

Gyro A. Madrona
Electronics Engineer

TOPIC OUTLINE

Shapiro-Wilk

Anderson-Darling

SHAPIRO-WILK

SHAPIRO-WILK

Shapiro-Wilk is a statistical test used to assess whether a given dataset follows a **normal distribution**. It is particularly effective for sample size $\boldsymbol{n \leq 50}$.

Hypothesis

$\boldsymbol{H_o}$: Normal data

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

Assumption

- Continuous data

syntax

```
from scipy import stats

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

Let $\alpha = 0.05$

Hypothesis

$\boldsymbol{H_o}$: Normal data

$\boldsymbol{H_a}$: Non-normal data (p-value ≤ 0.05)

ANDERSON-DARLING

ANDERSON-DARLING

Anderson-Darling is a statistical test used to assess whether a given dataset follows a **normal distribution**. It is particularly effective for sample size $\boldsymbol{n > 50}$.

Hypothesis

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

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

Assumption

- Continuous data

syntax

```
from scipy import stats

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

Let $\alpha = 0.05$

Hypothesis

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

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

LABORATORY