NORMALITY TEST

HYPOTHESIS TESTING

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 ($\boldsymbol{n \leq 50}$).

Hypothesis

$\boldsymbol{H_o}$: Normal data

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

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-30-sample.csv**"

Solution

```
from scipy import stats

w_stat, p_value =

stats.shapiro(df['Defects'])
```

Let $\alpha = 0.05$

Null Hypothesis

$\boldsymbol{H_o}$: Normal data

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

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 ($\boldsymbol{n > 50}$).

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, alpha =
    stats.anderson(data)
```

EXERCISE

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

dataset

"**defects-count.csv**"

Solution

```
from scipy import stats
a2_stat, critical, alpha =
    stats.anderson(df['Defects'])
```

Let $\alpha = 0.05$

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)

LABORATORY