T-TEST

HYPOTHESIS TESTING

prepared by:

Gyro A. Madrona
Electronics Engineer

TOPIC OUTLINE

1-Sample t-Test

F-Test

2-Sample t-Test

Paired t-Test

1-SAMPLE T-TEST

1-SAMPLE T-TEST

1-sample t-test is a statistical method used to determine whether the **mean** of a **single sample** ($\bar{x}_1$) differs significantly from a known or hypothesized population mean (μ_o).

Test Statistic

$$t = \frac{\bar{x}_1 - \mu_o}{s/\sqrt{n}}$$

where:

t = t-statistic

$\bar{x}_1$ = sample mean

μ_o = hypothesized population mean

s = sample standard deviation

n = sample size

1-SAMPLE T-TEST

syntax

```
from scipy import stats

t_stat, p_value = stats.ttest_1samp(
    sample_data, pop_mean,
    alternative = 'two-sided'
)
```

Null Hypothesis

$$H_o: \mu_1 = \mu_o$$

Alternative Hypothesis

$$H_a: \mu_1 \neq \mu_o \text{ (p-value} \leq \alpha)$$

Assumptions

- Continuous data
- Normal data

EXERCISE

The dataset contains the electricity production in MWh by the following production types:

Type	μ	σ
Nuclear	1282.48	241.19
Wind	771.73	666.53
Hydroelectric	1792.64	679.98
Oil and Gas	1166.84	433.92
Coal	1138.50	329.93
Solar	171.81	257.26
Biomass	54.94	13.76

dataset

"**electricity-sample-1-dataset.csv**"

Perform a **1-sample t-test** to determine whether the mean electricity production (in MWh) for each production type in the given dataset differs significantly from known population parameters.

F-TEST

F-TEST

F-test is a statistical test used to compare the **variances** of **two samples** (s_1^2, s_2^2) and determine if they are significantly different.

Test Statistic

$$F = s_1^2 / s_2^2$$

where:

F = F-statistic

s_1^2 = larger sample variance

s_2^2 = smaller sample variance

CUMULATIVE PROBABILITY OF F-DISTRIBUTION

syntax

```
from scipy import stats

# F-statistic

if var_1 > var_2:

    f_stat = var_1/var_2

else:

    f_stat = var_2/var_1


p_value = 1 - stats.f.cdf(

    f_stat,dof_1,dof_2

)
```

Null Hypothesis

$H_o: \sigma_1^2 = \sigma_2^2$

Alternative Hypothesis

$H_a: \sigma_1^2 \neq \sigma_2^2$ (p-value $\leq \alpha$)

Assumptions

- Continuous data
- Normal data

2-SAMPLE T-TEST

2-SAMPLE T-TEST

2-sample t-test is a statistical method used to compare the **means** of **two independent groups** $(\bar{x}_1, \bar{x}_2)$ to determine if they are significantly different from each other.

Test Statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where:

t = t-statistic

$\bar{x}_1, \bar{x}_2$ = mean of each group

n_1, n_2 = sample size of each group

s_P = pooled standard deviation

$$= \left(\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \right)$$

INDEPENDENT SAMPLES

syntax

```
from scipy import stats

t_stat, p_value = stats.ttest_ind(
    sample_1 data, sample_2 data,
    alternative = 'two-sided',
    equal_var = True
)
```

Null Hypothesis

$$H_o: \mu_1 = \mu_1$$

Alternative Hypothesis

$$H_a: \mu_1 \neq \mu_2 \quad (\text{p-value} \leq \alpha)$$

Assumptions

- Continuous data
- Normal data

PAIRED T-TEST

PAIRED T-TEST

Paired t-test is a statistical method used to compare the **means** of **two related groups** ($\bar{x}_1, \bar{x}_2$) to determine if they are significantly different from each other.

Test Statistic

$$t = \frac{\bar{x}_d}{s_d\sqrt{n}}$$

where:

t = t-statistic

$\bar{x}_d$ = mean of the difference between groups

s_d = standard deviation of the difference

n = sample size

RELATED SAMPLES

syntax

```
from scipy import stats

t_stat, p_value = stats.ttest_rel(
    sample_1 data, sample_2 data,
    alternative = 'two-sided',
    equal_var = True
)
```

Null Hypothesis

$$H_o: \mu_d = 0$$

Alternative Hypothesis

$$H_a: \mu_d \neq 0 \text{ (p-value} \leq \alpha)$$

Assumptions

- Continuous data
- Normal data

EXERCISE

The dataset contains the electricity production in MWh by the following production types:

Type	μ	σ
Nuclear	1282.48	241.19
Wind	771.73	666.53
Hydroelectric	1792.64	679.98
Oil and Gas	1166.84	433.92
Coal	1138.50	329.93
Solar	171.81	257.26
Biomass	54.94	13.76

dataset

"**electricity-sample-1-dataset.csv**"

Perform a **2-sample t-test** to compare the means of electricity production (in MWh) between two distinct groups in the dataset.

LABORATORY