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** differs significantly from a known or hypothesized population mean.

Hypothesis

$H_o: \mu_1 = \mu_o$

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

Assumptions

- Continuous data
- Normal data

syntax

```
from scipy import stats
t_stat, p_value = stats.ttest_1samp(
    sample_data, pop_mean)
```

options

```
alternative = 'two-sided' # default
alternative = 'larger'  # u1 > u0
alternative = 'smaller' # u1 < u0
```

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** and determine if they are significantly different.

Hypothesis

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

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

Assumptions

- Continuous data
- Normal data

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)
```

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** to determine if they are significantly different from each other.

Hypothesis

$H_o: \mu_1 = \mu_1$

$H_a: \mu_1 \neq \mu_2$ (p-value $\leq \alpha$)

Assumptions

- Continuous data
- Normal data

syntax

```
from scipy import stats
t_stat, p_value = stats.ttest_ind(
    sample_1 data, sample_2 data)
```

options

```
alternative = 'two-sided' # default
alternative = 'larger'  # u1 > u1
alternative = 'smaller' # u1 < u2
equal_var = True # default
```

PAIRED T-TEST

PAIRED T-TEST

Paired t-test is a statistical method used to compare the means of **two related groups** to determine if they are significantly different from each other.

Hypothesis

$H_o: \mu_1 = \mu_1$

$H_a: \mu_1 \neq \mu_2$ (p-value $\leq \alpha$)

Assumptions

- Continuous data
- Normal data

syntax

```
from scipy import stats
t_stat, p_value = stats.ttest_rel(
    sample_1 data, sample_2 data)
```

options

```
alternative = 'two-sided' # default
alternative = 'larger'  # u1 > u1
alternative = 'smaller' # u1 < u2
equal_var = True # default
```

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