Analysis if Variance:

How far a set of information varies

1. 1 way Classification

Using one variable for analysis

Mobile sales rejection rate.in percentage%

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Mobile Brands | Chennai | Delhi | Bangalore | Kolkata | Mumbai | Madurai |
| Nokia | 24 | 42 | 60 | 50 | 60 | 42 |
| Samsung | 10 | 15 | 15 | 24 | 15 | 15 |
| Oppo | 20 | 25 | 40 | 10 | 40 | 25 |
| Apple | 5 | 10 | 30 | 30 | 30 | 10 |
| Jio | 50 | 5 | 10 | 50 | 10 | 5 |
| Vivo | 40 | 49 | 45 | 40 | 45 | 49 |

H0 – Null Hypothesis – There is no difference in the mobile brands rejections

H1 – Alternate Hypothesis – There is difference in the mobile brand rejections

if the result of the calculations is true, H0 is true and H1 is false. Else H1 is true and H0 is false.

1. 2 way Classification
   1. Using 2 variables for analysis
   2. TV Models and Cost in $ value in different Store / Shops

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Models** | **Deal one** | **V&Co$** | **Darling$** | **S- Stores** |
| Samsung | 14000 | 34000 | 18000 | 250000 |
| Onida | 11000 | 12000 | 19000 | 110000 |
| Smart TV | 12000 | 21000 | 14000 | 150000 |
| Vandio | 24000 | 42000 | 45000 | 100000 |

1. Cost of Mobiles in each city INR

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Mobile Brands | Chennai | Delhi | Bangalore | Kolkata | Mumbai | Madurai |
| Nokia | 24K | 42 K | 60 K | 50K | 60 K | 42 K |
| Samsung | 10 K | 15 K | 15 K | 24K | 15 K | 15 K |
| Oppo | 20 K | 25 K | 40 K | 10 K | 40 K | 25 K |
| Apple | 5 K | 10 K | 30 K | 30 K | 30 K | 10 K |
| Jio | 50 K | 55 K | 10 K | 50 K | 10 K | 55 K |
| Vivo | 40 K | 49 K | 45 K | 40 K | 45 K | 49 K |

H0 – there is significant difference between the brands cost

H1- there is no significant difference between the brands cost

1. Placement data
   1. 12th,10th Degree pass marks – one way classification
   2. Pass marks with Students wise - Two way Classification
2. Metals values over the years. In Sovereign and Price in Thousands INR

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Metals | 2005 | 2010 | 2015 | 2020 | 2025 |
| Gold | 24 | 42 | 60 | 50 | 60 |
| Silver | 10 | 15 | 15 | 24 | 15 |
| Platinum | 20 | 25 | 40 | 10 | 40 |

Yearly price value of the precious metals in India.

**Two-Way ANOVA: Examples**

Various examples where two way ANOVA concepts are used:

* **Medicine Experiment**: Testing the effect of two types of medicine (A and B) on patients from different age groups (young and old).
* **Crop Yield Study**: Analyzing the impact of two fertilizers (X and Y) on crop yield across different soil types (sandy and loamy).
* **Education Intervention**: Evaluating the effectiveness of two teaching methods (traditional and online) on students from different socioeconomic backgrounds (low-income and high-income).
* **Marketing Campaign**: Assessing the influence of two advertising strategies (social media and television) on customer response among different geographical regions (urban and rural).
* **Fitness Program Evaluation**: Investigating the effects of two exercise routines (aerobic and strength training) on fitness levels among various age categories (teens, adults, seniors).

Sample:

import pandas as pd

import statsmodels.api as sm

from statsmodels.formula.api import ols

# Example DataFrame

data = pd.DataFrame({

'exercise\_type': ['Aerobic', 'Aerobic', 'Strength', 'Strength', 'Aerobic', 'Strength'] \* 10,

'age\_group': ['Teen', 'Adult', 'Senior', 'Teen', 'Adult', 'Senior'] \* 10,

'fitness\_change': [4.2, 5.1, 3.5, 4.8, 5.0, 3.3] \* 10

})

# Two-way ANOVA model

model = ols('fitness\_change ~ C(exercise\_type) + C(age\_group) + C(exercise\_type):C(age\_group)', data=data).fit()

anova\_table = sm.stats.anova\_lm(model, typ=2)

print(anova\_table)

**Dependent variable (DV):**  
→ fitness\_score or fitness\_change (post - pre)

**Independent variables (factors):**

* exercise\_type → categorical (2 levels: Aerobic, Strength)
* age\_group → categorical (3 levels: Teen, Adult, Senior)

**Hypotheses**

**Main effects**

* H₀₁: No significant difference in fitness levels across exercise types.
* H₀₂: No significant difference in fitness levels across age groups.

| **Source** | **Sum Sq** | **df** | **F** | **PR(>F)** |
| --- | --- | --- | --- | --- |
| C(exercise\_type) | 12.35 | 1 | 5.67 | 0.021 |
| C(age\_group) | 18.42 | 2 | 4.23 | 0.018 |
| C(exercise\_type):C(age\_group) | 8.77 | 2 | 2.01 | 0.138 |
| Residual | 130.24 | 54 |  |  |

P value is < 0.05 Null Hypothesis is accepted.

Hence Alternate Hypothesis is rejected.