ANOVA (Analysis of Variance) and Chi-Square Test

# ANOVA (Analysis of Variance)

## Overview

ANOVA (Analysis of Variance) is a statistical method used to compare the means of three or more samples to determine if at least one sample mean is different from the others. It helps in identifying whether the variations between group means are significant.

## Types of ANOVA

### One-Way ANOVA

Compares the means of three or more independent groups based on one factor.  
Example: Comparing the test scores of students from different classes.

### Two-Way ANOVA

Compares the means based on two factors and can also examine the interaction between the factors.  
Example: Comparing the effect of different teaching methods and different study times on test scores.

## Assumptions

Independence: The samples must be independent of each other.  
Normality: The data in each group should be approximately normally distributed.  
Homogeneity of Variances: The variances among the groups should be equal.

## Hypotheses

Null Hypothesis (H₀): All group means are equal.  
Alternative Hypothesis (H₁): At least one group mean is different.

## Calculation

Between-Group Variance:  
Measures the variation due to the interaction between the groups.  
Calculated as the sum of the squared differences between the group means and the overall mean.  
  
Within-Group Variance:  
Measures the variation within each group.  
Calculated as the sum of the squared differences within each group.  
  
F-Statistic:  
The ratio of the between-group variance to the within-group variance.  
If the F-statistic is significantly large, it suggests that at least one group mean is different.

## Example

One-Way ANOVA:  
```python  
import scipy.stats as stats  
  
# Example data  
group1 = [23, 25, 27, 22, 24]  
group2 = [35, 37, 36, 38, 39]  
group3 = [45, 47, 44, 46, 48]  
  
# Perform One-Way ANOVA  
f\_statistic, p\_value = stats.f\_oneway(group1, group2, group3)  
print("F-Statistic:", f\_statistic)  
print("P-Value:", p\_value)  
```

## Interpretation

If p-value < α (e.g., 0.05): Reject the null hypothesis, indicating that at least one group mean is significantly different.  
If p-value ≥ α: Fail to reject the null hypothesis, indicating no significant difference between the group means.

# Chi-Square Test

## Overview

The Chi-Square Test is a statistical method used to determine if there is a significant association between two categorical variables. It compares the observed frequencies in each category to the expected frequencies.

## Types of Chi-Square Tests

### Chi-Square Test of Independence

Determines if there is an association between two categorical variables.  
Example: Testing the relationship between gender and preference for a new product.

### Chi-Square Goodness of Fit Test

Determines if a sample matches the expected distribution.  
Example: Testing if a die is fair by comparing the observed roll frequencies to the expected frequencies.

## Assumptions

Independence: The observations must be independent of each other.  
Sample Size: Each category should have a sufficiently large sample size (expected frequency of at least 5).

## Hypotheses

Null Hypothesis (H₀): There is no association between the variables (for test of independence) or the observed distribution matches the expected distribution (for goodness of fit).  
Alternative Hypothesis (H₁): There is an association between the variables (for test of independence) or the observed distribution does not match the expected distribution (for goodness of fit).

## Calculation

Observed Frequencies (O):  
The actual count of observations in each category.  
  
Expected Frequencies (E):  
The count of observations expected in each category if the null hypothesis is true.  
Calculated based on the total sample size and the marginal totals for each category.  
  
Chi-Square Statistic (χ²):  
The sum of the squared differences between the observed and expected frequencies, divided by the expected frequencies.  
χ² = Σ [(O\_i - E\_i)² / E\_i]

## Example

Chi-Square Test of Independence:  
```python  
import scipy.stats as stats  
import numpy as np  
  
# Example data  
data = np.array([[10, 20], [20, 40]])  
  
# Perform Chi-Square Test  
chi2\_stat, p\_val, dof, ex = stats.chi2\_contingency(data)  
print("Chi-Square Statistic:", chi2\_stat)  
print("P-Value:", p\_val)  
```

## Interpretation

If p-value < α (e.g., 0.05): Reject the null hypothesis, indicating a significant association between the variables.  
If p-value ≥ α: Fail to reject the null hypothesis, indicating no significant association between the variables.

## Applications

Market Research: Analyzing customer preferences and behavior.  
Medical Research: Studying the relationship between risk factors and health outcomes.  
Social Sciences: Investigating the association between demographic variables and attitudes.

ANOVA (Analysis of Variance) and Chi-Square Test

# ANOVA (Analysis of Variance)

## Overview

ANOVA (Analysis of Variance) is a statistical method used to compare the means of three or more samples to determine if at least one sample mean is different from the others. It helps in identifying whether the variations between group means are significant.

## Types of ANOVA

### One-Way ANOVA

Compares the means of three or more independent groups based on one factor.  
Example: Comparing the test scores of students from different classes.

### Two-Way ANOVA

Compares the means based on two factors and can also examine the interaction between the factors.  
Example: Comparing the effect of different teaching methods and different study times on test scores.

## Assumptions

Independence: The samples must be independent of each other.  
Normality: The data in each group should be approximately normally distributed.  
Homogeneity of Variances: The variances among the groups should be equal.

## Hypotheses

Null Hypothesis (H₀): All group means are equal.  
Alternative Hypothesis (H₁): At least one group mean is different.

## Calculation

### Between-Group Variance

Measures the variation due to the interaction between the groups.  
Calculated as the sum of the squared differences between the group means and the overall mean.

### Within-Group Variance

Measures the variation within each group.  
Calculated as the sum of the squared differences within each group.

### F-Statistic

The ratio of the between-group variance to the within-group variance.  
If the F-statistic is significantly large, it suggests that at least one group mean is different.

## Example

One-Way ANOVA:  
```python  
import scipy.stats as stats  
  
# Example data  
group1 = [23, 25, 27, 22, 24]  
group2 = [35, 37, 36, 38, 39]  
group3 = [45, 47, 44, 46, 48]  
  
# Perform One-Way ANOVA  
f\_statistic, p\_value = stats.f\_oneway(group1, group2, group3)  
print("F-Statistic:", f\_statistic)  
print("P-Value:", p\_value)  
```

## Interpretation

If p-value < α (e.g., 0.05): Reject the null hypothesis, indicating that at least one group mean is significantly different.  
If p-value ≥ α: Fail to reject the null hypothesis, indicating no significant difference between the group means.

# Chi-Square Test

## Overview

The Chi-Square Test is a statistical method used to determine if there is a significant association between two categorical variables. It compares the observed frequencies in each category to the expected frequencies.

## Types of Chi-Square Tests

### Chi-Square Test of Independence

Determines if there is an association between two categorical variables.  
Example: Testing the relationship between gender and preference for a new product.

### Chi-Square Goodness of Fit Test

Determines if a sample matches the expected distribution.  
Example: Testing if a die is fair by comparing the observed roll frequencies to the expected frequencies.

## Assumptions

Independence: The observations must be independent of each other.  
Sample Size: Each category should have a sufficiently large sample size (expected frequency of at least 5).

## Hypotheses

Null Hypothesis (H₀): There is no association between the variables (for test of independence) or the observed distribution matches the expected distribution (for goodness of fit).  
Alternative Hypothesis (H₁): There is an association between the variables (for test of independence) or the observed distribution does not match the expected distribution (for goodness of fit).

## Calculation

### Observed Frequencies (O)

The actual count of observations in each category.

### Expected Frequencies (E)

The count of observations expected in each category if the null hypothesis is true.  
Calculated based on the total sample size and the marginal totals for each category.

### Chi-Square Statistic (χ²)

The sum of the squared differences between the observed and expected frequencies, divided by the expected frequencies.  
χ² = Σ [(O\_i - E\_i)² / E\_i]

## Example

Chi-Square Test of Independence:  
```python  
import scipy.stats as stats  
import numpy as np  
  
# Example data  
data = np.array([[10, 20], [20, 40]])  
  
# Perform Chi-Square Test  
chi2\_stat, p\_val, dof, ex = stats.chi2\_contingency(data)  
print("Chi-Square Statistic:", chi2\_stat)  
print("P-Value:", p\_val)  
```

## Interpretation

If p-value < α (e.g., 0.05): Reject the null hypothesis, indicating a significant association between the variables.  
If p-value ≥ α: Fail to reject the null hypothesis, indicating no significant association between the variables.

## Applications

Market Research: Analyzing customer preferences and behavior.  
Medical Research: Studying the relationship between risk factors and health outcomes.  
Social Sciences: Investigating the association between demographic variables and attitudes.