

# Chi-Square Test

## Example 1: Fair Die Test

A fair die is rolled 120 times and the following results are obtained:

- Face 1: 22 times
- Face 2: 17 times
- Face 3: 20 times
- Face 4: 26 times
- Face 5: 22 times
- Face 6: 13 times

Test at a 5% level of significance whether the die is fair.

## Define Hypotheses

Null Hypothesis: The die is fair

Alternative Hypothesis: The die is biased

## Import libraries

```
import scipy.stats as st
import numpy as np

# Observed frequencies
observed = np.array([22, 17, 20, 26, 22, 13])

# Expected frequencies (since the die is fair, each face should appear 120/6 times)
expected = np.array([20, 20, 20, 20, 20, 20])
# Set significance level
alpha = 0.05

# Degrees of freedom (df = categories - 1)
df = len(observed) - 1

# Perform Chi-Square test
chi_sq = np.sum(((observed-expected)**2)/expected)
# Print results
print(f"Chi-Square Value:{chi_sq:.2f}")
```

Chi-Square Value:5.10

```

#decision making using critical value
critical_value = st.chi2.ppf(1-alpha,df)
print(f"critical value:{critical_value:.2f}")

critical value:11.07

#compare:
if chi_sq > critical_value:
    print("null hypothesis is rejected: the dice isn't fair")
else:
    print("fail to reject null hypothesis: the dice is fair")

fail to reject null hypothesis: the dice is fair

```

## Example 2: Gender and Music Preference

A study was conducted to investigate whether there is a relationship between gender and the preferred genre of music. A sample of 235 people was selected, and the data collected is shown below. Test at a 5% level of significance whether there is a significant association between gender and music preference.

Genre	Male	Female
Pop	40	35
Hip Hop	45	30
Classical	25	20
Rock	10	30

## Define Hypotheses

Null Hypothesis: There is no association between gender and music preference (they are independent).

Alternative Hypothesis: There is an association between gender and music preference.

```

observed = np.array([[40,45,25,10],
                    [35,30,20,30]])

#perform chi-square test
chi2_stat,p_value,df,expected=st.chi2_contingency(observed)
print(f'chi-square:{chi2_stat}')
print(f'degree of freedom:{df}')
print(f'expected freq:{expected}')
print(f"p-value:{p_value}")

```

```
chi-square:13.788747987117553
degree of freedom:3
expected freq:[[38.29787234 38.29787234 22.9787234 20.42553191]
 [36.70212766 36.70212766 22.0212766 19.57446809]]
p-value:0.0032072711944191893
```

```
#finding critical value
```

```
critical_value = st.chi.ppf(1-0.05,df)
print(f"critical_value:{critical_value}")
```

```
critical_value:2.7954834829151074
```

```
#Take decision
```

```
if chi2_stat>critical_value:
    print("Reject the null hypothesis:")
else:
    print("fail to reject null hypothesis")
```

```
#based on p-value
```

```
if p_value<0.05:
    print("Reject the null hypothesis:")
else:
    print("fail to reject null hypothesis")
```

```
Reject the null hypothesis:
```

```
Reject the null hypothesis:
```