chi-square

November 1, 2024

[1]:
```
from scipy.stats import chi2_contingency # defining the table
data = [[207, 282, 241], [234, 242, 232]]
stat, p, dof, expected = chi2_contingency(data) # interpret p-value
alpha = 0.05
print("p value is " + str(p))
if p <= alpha:
    print('Dependent (reject H0)')
else:
    print('Independent (H0 holds true)')
```

```
p value is 0.10319714047309392
Independent (H0 holds true)
```

[2]:
```
import numpy as np
from scipy.stats import chi2

# Observed frequencies
observed = np.array([115, 47, 41, 101, 200, 96])

# Expected frequencies (assuming a fair die)
expected = np.array([100, 100, 100, 100, 100, 100])

# Calculate chi-square statistic
chi2_stat = np.sum((observed - expected)**2 / expected)

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

# Critical value for 10% significance level
critical_value = chi2.ppf(0.90, df)

# p-value
p_value = 1 - chi2.cdf(chi2_stat, df)

# Output results
print(f"Chi-squared Statistic: {chi2_stat}")
print(f"Critical Value at 10% significance level: {critical_value}")
print(f"p-value: {p_value}")
```

```
# Conclusion
if chi2_stat < critical_value:
    print("Fail to reject the null hypothesis: The die is unbiased.")
else:
    print("Reject the null hypothesis: The die is biased.")
```

```
Chi-squared Statistic: 165.32000000000002
Critical Value at 10% significance level: 9.236356899781123
p-value: 0.0
Reject the null hypothesis: The die is biased.
```

[3]:
```
import numpy as np
import pandas as pd
from scipy.stats import chi2_contingency

# Define the observed data
data = np.array([
    [10, 102, 8],   # Machine 1
    [34, 161, 5],   # Machine 2
    [12, 79, 9],    # Machine 3
    [10, 60, 10]    # Machine 4
])

# Create a DataFrame for better visualization (optional)
df = pd.DataFrame(data, columns=['Too Thin', 'OK', 'Too Thick'],
                  index=['Machine 1', 'Machine 2', 'Machine 3', 'Machine 4'])

print("Observed Data:\n", df)

# Perform the Chi-Square test
chi2_stat, p_value, dof, expected = chi2_contingency(data)

# Display results
print("\nChi-Square Statistic:", chi2_stat)
print("P-Value:", p_value)
print("Degrees of Freedom:", dof)
print("Expected Frequencies:\n", expected)

# Determine if the result is significant
alpha = 0.05
if chi2_stat > chi2.ppf(1 - alpha, dof):
    print("Reject the null hypothesis: There is a significant difference.")
else:
    print("Fail to reject the null hypothesis: No significant difference.")
```

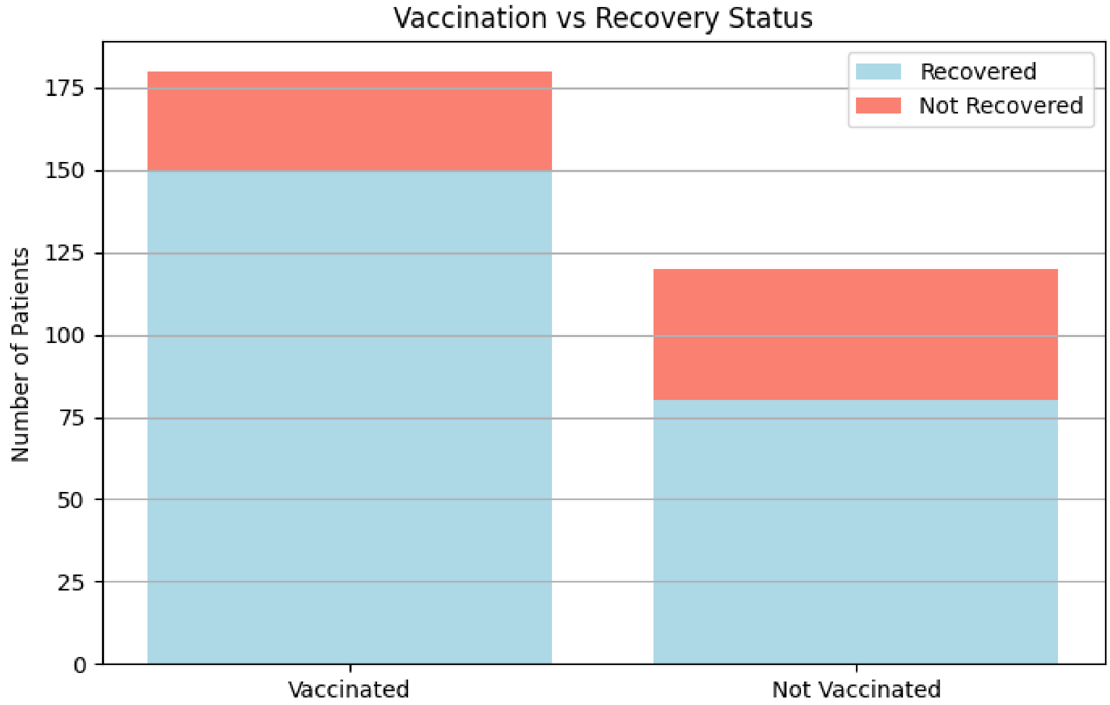
```
Observed Data:
            Too Thin   OK  Too Thick
```

```
Machine 1        10  102          8
Machine 2        34  161          5
Machine 3        12   79          9
Machine 4        10   60         10

Chi-Square Statistic: 15.584353328056686
P-Value: 0.01616760116149423
Degrees of Freedom: 6
Expected Frequencies:
 [[ 15.84  96.48   7.68]
 [ 26.4  160.8   12.8 ]
 [ 13.2   80.4    6.4 ]
 [ 10.56  64.32   5.12]]
Reject the null hypothesis: There is a significant difference.
```

[4]:
```python
import numpy as np
import pandas as pd
from scipy.stats import chi2_contingency
import matplotlib.pyplot as plt

# Create a contingency table
data = np.array([[150, 30],     # Vaccinated
                 [80, 40]])    # Not Vaccinated

# Display the contingency table as a DataFrame for clarity
contingency_table = pd.DataFrame(data,
                                  columns=['Recovered', 'Not Recovered'],
                                  index=['Vaccinated', 'Not Vaccinated'])
print("Contingency Table:\n", contingency_table)

# Perform the Chi-Square test
chi2_stat, p_value, dof, expected = chi2_contingency(data)

# Display results
print("\nChi-Square Statistic:", chi2_stat)
print("P-Value:", p_value)
print("Degrees of Freedom:", dof)
print("Expected Frequencies:\n", expected)

# Determine significance level
alpha = 0.05
if p_value < alpha:
    print("Reject the null hypothesis: There is a significant association␣
 ↪between vaccination and recovery.")
else:
    print("Fail to reject the null hypothesis: No significant association␣
 ↪between vaccination and recovery.")
```

```python
# Optional: Plotting the contingency table
plt.figure(figsize=(8, 5))
plt.title("Vaccination vs Recovery Status")
plt.bar(['Vaccinated', 'Not Vaccinated'], [150, 80], label='Recovered',␣
 ↪color='lightblue')
plt.bar(['Vaccinated', 'Not Vaccinated'], [30, 40], label='Not Recovered',␣
 ↪color='salmon', bottom=[150, 80])
plt.ylabel('Number of Patients')
plt.legend()
plt.grid(axis='y')
plt.show()
```

```
Contingency Table:
                 Recovered  Not Recovered
Vaccinated             150             30
Not Vaccinated          80             40

Chi-Square Statistic: 10.267857142857142
P-Value: 0.0013536793727780064
Degrees of Freedom: 1
Expected Frequencies:
 [[138.  42.]
 [ 92.  28.]]
Reject the null hypothesis: There is a significant association between
vaccination and recovery.
```



[5]:
```python
import numpy as np
import pandas as pd
from scipy.stats import chi2_contingency
import matplotlib.pyplot as plt

# Create a contingency table
data = np.array([[30, 10],  # Male
                 [20, 30]]) # Female

# Display the contingency table as a DataFrame for clarity
contingency_table = pd.DataFrame(data,
                                  columns=['Purchased', 'Not Purchased'],
                                  index=['Male', 'Female'])
print("Contingency Table:\n", contingency_table)

# Perform the Chi-Square test
chi2_stat, p_value, dof, expected = chi2_contingency(data)

# Display results
print("\nChi-Square Statistic:", chi2_stat)
print("P-Value:", p_value)
print("Degrees of Freedom:", dof)
print("Expected Frequencies:\n", expected)

# Determine significance level
alpha = 0.05
if p_value < alpha:
    print("Reject the null hypothesis: There is a significant association␣
 ↪between gender and product preference.")
else:
    print("Fail to reject the null hypothesis: No significant association␣
 ↪between gender and product preference.")

# Optional: Plotting the contingency table
plt.figure(figsize=(8, 5))
plt.title("Gender vs Product Purchase Preference")
plt.bar(['Male', 'Female'], [30, 20], label='Purchased', color='lightblue')
plt.bar(['Male', 'Female'], [10, 30], label='Not Purchased', color='salmon',␣
 ↪bottom=[30, 20])
plt.ylabel('Number of Individuals')
plt.legend()
plt.grid(axis='y')
plt.show()
```

```
Contingency Table:
```

```
         Purchased  Not Purchased
Male            30             10
Female          20             30

Chi-Square Statistic: 9.6530625
P-Value: 0.001890361677058677
Degrees of Freedom: 1
Expected Frequencies:
 [[22.22222222 17.77777778]
 [27.77777778 22.22222222]]
Reject the null hypothesis: There is a significant association between gender
and product preference.
```