Multi category chi-squared tests: Takeaways

Syntax

- Calculating the chi-squared value:

```python
observed = [6662, 1179, 15128, 9592]

expected = [5257.6, 2589.6, 16558.2, 8155.6]

values = []

for i, obs in enumerate(observed):

    exp = expected[i]

    value = (obs - exp) ** 2 / exp

    values.append(value)

chisq_gender_income = sum(values)
```

- Finding the chi-squared value and p-value using scipy.stats.chisquare:

```python
import numpy as np

from scipy.stats import chisquare

observed = np.array([6662, 1179, 15128, 9592])

expected = np.array([[5257.6, 2589.6, 16558.2, 8155.6])

chisq_value, pvalue_gender_income = chisquare(observed, expected)
```

- Using the pandas.crosstab function to print a table that shows frequency counts:

```python
import pandas

table = pandas.crosstab(income["sex"], [income["high_income"]])

print(table)
```

- Using the scipy.stats.chi2_contingency function to generate the expected values:

```python
import numpy as np

from scipy.stats import chi2_contingency

observed = np.array([[5, 5], [10, 10]])

chisq_value, pvalue, df, expected = chi2_contingency(observed)
```

Concepts

- In a multiple category chi-squared test, we calculate expected values across our whole dataset.
- We can calculate the chi-squared value by using the following steps:
  - Subtract the expected value from the observed value.
  - Subtract the difference.
  - Divide the squared difference by the expected value.
  - Repeat for all observed and expected values and add up all the values.
- Formula for chi-squared:

$$\sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

- Finding that a result isn't significant doesn't mean that no association between the columns exists. Finding a statistically significant result doesn't imply anything about what the correlation is.
- Chi-squared tests can only be applied in the case where each possibility within a category is independent.

Resources

- [Chi-squared test of association](#)
- [Documentation for scipy.stats.chi2_contingency function](#)
- [Documentation for pandas.crosstab function](#)