Hypothesis Testing

November 20, 2023

```
[37]: import numpy as np
      from scipy.stats import chi2_contingency
      import scipy.stats as stats
```

1 Alcohol and Substance Abuse: Chi Squared Tests (Income, Region, Race, Gender)

1.1 Race

```
[8]: # Create a contingency table
     observed_data = np.array([[142,23890],
                               [146,118733],
                               [383,176839],
                               [1001,474737],
                               [1130,581671],
                               [4004,1314526]])
```

```
[15]: # Perform the chi-squared test
      chi2, p, dof, expected = chi2_contingency(observed_data)
```

```
[17]: chi2
```

```
[17]: 448.2708136665917
```

```
[18]: p
```

```
[18]: 1.1594774961384505e-94
```

```
[12]: # Check the p-value to determine statistical significance
      alpha = 0.05  # Set your chosen significance level
      if p < alpha:
          print("Reject the null hypothesis: The number of alcohol abuse cases is␣
       ↪dependent on race.")
      else:
          print("Fail to reject the null hypothesis: The number of alcohol abuse␣
       ↪cases is independent of race.")
```

```
Reject the null hypothesis: The number of alcohol abuse cases is dependent on
race.
```

1.2 Gender

```
[19]: # Create a contingency table
      observed_data = np.array([[4488,1354413],
                                [2708,1544041],
                                ])
```

```
[20]: # Perform the chi-squared test
      chi2, p, dof, expected = chi2_contingency(observed_data)
```

```
[22]: chi2
```

```
[22]: 704.5858230765883
```

```
[23]: p
```

```
[23]: 3.0094796136033507e-155
```

```
[21]: # Check the p-value to determine statistical significance
      alpha = 0.05  # Set your chosen significance level
      if p < alpha:
          print("Reject the null hypothesis: The number of alcohol abuse cases is␣
       ↪dependent on gender.")
      else:
          print("Fail to reject the null hypothesis: The number of alcohol abuse␣
       ↪cases is independent of gender.")
```

```
Reject the null hypothesis: The number of alcohol abuse cases is dependent on
gender.
```

1.3 Region

```
[27]: # Create a contingency table
      observed_data = np.array([[1467,662619],
                                [2602,1172837],
                                [1257,388398],
                                [1871,674599]])
```

```
[28]: # Perform the chi-squared test
      chi2, p, dof, expected = chi2_contingency(observed_data)
```

```
[29]: chi2
```

```
[29]: 163.5906715202513
```

[30]:
```
p
```

[30]:
```
3.077423050012854e-35
```

[31]:
```
# Check the p-value to determine statistical significance
alpha = 0.05  # Set your chosen significance level
if p < alpha:
    print("Reject the null hypothesis: The number of alcohol abuse cases is␣
 ↪dependent on region.")
else:
    print("Fail to reject the null hypothesis: The number of alcohol abuse␣
 ↪cases is independent of region.")
```

```
Reject the null hypothesis: The number of alcohol abuse cases is dependent on
region.
```

1.4 Income

[32]:
```
# Create a contingency table
observed_data = np.array([[1540,557327],
                          [1918,777093],
                          [1690,678509],
                          [2049,885524]])
```

[33]:
```
# Perform the chi-squared test
chi2, p, dof, expected = chi2_contingency(observed_data)
```

[34]:
```
chi2
```

[34]:
```
27.833423644850136
```

[35]:
```
p
```

[35]:
```
3.936499874097809e-06
```

[36]:
```
# Check the p-value to determine statistical significance
alpha = 0.05  # Set your chosen significance level
if p < alpha:
    print("Reject the null hypothesis: The number of alcohol abuse cases is␣
 ↪dependent on income.")
else:
    print("Fail to reject the null hypothesis: The number of alcohol abuse␣
 ↪cases is independent of income.")
```

```
Reject the null hypothesis: The number of alcohol abuse cases is dependent on
income.
```

1.5 Age : Correlation Test

[43]:

```
countofasa_age =␣
 ↪[296,24,19,18,13,15,14,17,17,33,20,23,65,84,181,370,691,707,1086,1739,1765]

# Calculate the Pearson correlation coefficient
correlation_coefficient, p_value = stats.pearsonr(age_alcohol_abuse,␣
 ↪countofasa_age)

# Output the results
print(f"Pearson correlation coefficient: {correlation_coefficient:.2f}")
print(f"P-value: {p_value:.2f}")

# Interpret the results
if p_value < 0.05:  # You can choose your significance level
    print("There is a significant correlation between age and the count of␣
 ↪cases.")
else:
    print("There is no significant correlation between age and the count of␣
 ↪cases.")
```

```
Pearson correlation coefficient: 0.73
P-value: 0.00
There is a significant correlation between age and the count of cases.
```

[ ]:

```
```