statistics-lab7-assignment

November 25, 2023

Q.1 A researcher gave an aptitude test to 24 respondents, 12 were men and 12 of them were women. He recorded the scores for each of the responded and tabulated it in the table below:

Men 80 79 92 65 83 84 95 78 81 85 73 52

Women 82 87 89 91 93 76 74 70 88 99 61 94

Use this data provided to test the null hypothesis that the distribution of scores is the same for men as for women. Use a significance level of 0.05. (Perform U-test)

[8]:
```
import numpy as np

men_scores = [80, 79, 92, 65, 83, 84, 95, 78, 81, 85, 73, 52]
women_scores = [82, 87, 89, 91, 93, 76, 74, 70, 88, 99, 61, 94]

all_scores = np.concatenate([men_scores, women_scores])
ranks = np.argsort(np.argsort(all_scores)) + 1

U_men = np.sum(ranks[:len(men_scores)])
U_women = np.sum(ranks[len(men_scores):])

U = min(U_men, U_women)

n1 = len(men_scores)
n2 = len(women_scores)
critical_val = n1 * n2

if U >= critical_val:
    p_value = 2 * (1 - (U / (n1 * n2)))
else:
    p_value = 2 * (U / (n1 * n2))

alpha = 0.05

print("Test Statistic (U): ", U)
print("P-value: ", p_value)

if p_value < alpha:
```

```
    print("Reject the null hypothesis. There is significant evidence that the␣
 ↪distributions of scores for men and women are different.")
else:
    print("Fail to reject the null hypothesis. There is no significant evidence␣
 ↪that the distributions of scores for men and women are different.")
```

```
Test Statistic (U):  132
P-value:  1.8333333333333333
There is no significant evidence that the distributions of scores for men and
women are different.
```

Q.2 Three different kinds of food are tested on three groups of rats for 5 weeks. The objective is to check the difference in mean weight(in grams) of the rats per week. Apply one-way ANOVA using a 0.05 significance level to the following data:

Food I Food II Food III

8 4 11

12 5 8

19 4 7

8 6 13

6 9 7

11 7 9

[9]:
```
import numpy as np
from scipy.stats import f_oneway

food_I = np.array([8, 12, 19, 8, 6, 11])
food_II = np.array([4, 5, 4, 6, 9, 7])
food_III = np.array([11, 8, 7, 13, 7, 9])

f_statistic, p_value = f_oneway(food_I, food_II, food_III)

alpha = 0.05

print("F-statistic: ", f_statistic)
print(f"P-value: ",p_value)

if p_value < alpha:
    print("Reject the null hypothesis. There is significant evidence of at␣
 ↪least one mean difference between the groups.")
else:
    print("Fail to reject the null hypothesis. There is no significant evidence␣
 ↪of mean differences between the groups.")
```

```
F-statistic:  3.5537634408602146
P-value:  0.054529715225847826
```

```
There is no significant evidence of mean differences between the groups.
```

Q3. A researcher recruits 30 students to participate in a study. The students are randomly assigned to use one of three studying techniques for the next three weeks to prepare for an exam. At the end of the three weeks, all of the students take the same test. Use the following steps to perform a one-way ANOVA to determine if the average scores are the same across all three groups. (f_oneway() function from the SciPy library) group1 = [85, 86, 88, 75, 78, 94, 98, 79, 71, 80] group2 = [91, 92, 93, 85, 87, 84, 82, 88, 95, 96] group3 = [79, 78, 88, 94, 92, 85, 83, 85, 82, 81]

[10]:
```
from scipy.stats import f_oneway

group1 = [85, 86, 88, 75, 78, 94, 98, 79, 71, 80]
group2 = [91, 92, 93, 85, 87, 84, 82, 88, 95, 96]
group3 = [79, 78, 88, 94, 92, 85, 83, 85, 82, 81]

f_statistic, p_value = f_oneway(group1, group2, group3)

alpha = 0.05

print("F-statistic: ",f_statistic)
print("P-value: ",p_value)

if p_value < alpha:
    print("Reject the null hypothesis. There is significant evidence that the␣
 ↪average scores are different across the three groups.")
else:
    print("Fail to reject the null hypothesis. There is no significant evidence␣
 ↪that the average scores are different across the three groups.")
```

```
F-statistic:  2.3575322551335636
P-value:  0.11384795345837218
Fail to reject the null hypothesis. There is no significant evidence that the
average scores are different across the three groups.
```

Q4. We'll look at a table that illustrates the number of men and women who have purchased various types of fruits.

Apple Mango Banana Strawberry women 203 150 190 305 men 195 170 250 400 sum 398 320 440 705 Fruits buying table

The test's goal is to determine whether the two variables gender and fruit preference are connected. Make null hypothesis. Do using python code. (Use chi2_contingency()from scipy.stats).

[11]:
```
from scipy.stats import chi2_contingency

observed = [
    [203, 150, 190, 305],
    [195, 170, 250, 400] ]

chi2_stat, p_val, dof, expected = chi2_contingency(observed)
```

```
alpha = 0.05

print("Chi-squared Statistic: ",chi2_stat)
print("P-value: ",p_val)

if p_val < alpha:
    print("Reject the null hypothesis. There is evidence of an association␣
 ↪between gender and fruit preference.")
else:
    print("Fail to reject the null hypothesis. There is no evidence of an␣
 ↪association between gender and fruit preference.")
```

```
Chi-squared Statistic:  7.4842384990841575
P-value:  0.057964818341161385
Fail to reject the null hypothesis. There is no evidence of an association
between gender and fruit preference.
```

Q5. Table that shows the number of men and women buying different types of pets.

dog cat bird total

men 207 282 241 730

women 234 242 232 708

total 441 524 473 1438

State the null hypothesis for getting the relationship of variables. Calculate degrees of freedom.

[12]:
```
from scipy.stats import chi2_contingency

observed = [
    [207, 282, 241],
    [234, 242, 232]]

chi2_stat, p_val, dof, expected = chi2_contingency(observed)

alpha = 0.05

print(f"Chi-squared Statistic: ",chi2_stat)
print(f"P-value: ",p_val)
print(f"Degrees of Freedom: ",dof)

if p_val < alpha:
    print("Reject the null hypothesis. There is evidence of an association␣
 ↪between gender and type of pet purchased.")
else:
    print("Fail to reject the null hypothesis. There is no evidence of an␣
 ↪association between gender and type of pet purchased.")
```

```
Chi-squared Statistic:  4.542228269825232
P-value:  0.1031971404730939
Degrees of Freedom:  2
Fail to reject the null hypothesis. There is no evidence of an association
between gender and type of pet purchased.
```

Q6. A clinic provides a program to help their clients lose weight and asks a consumer agency to investigate the effectiveness of the program. The agency takes a sample of 15 people, weighing each person in the sample before the program begins and 3 months later. The results a tabulated below:

Determine the program is effective. Use python code. (Use stats.ttest_rel(before, after) from scipy.stats)

[13]:
```
from scipy import stats

before = [210, 205, 193, 182, 259, 239, 164, 197, 222, 211, 187, 175, 186, 243,␣
 ↪246]
after = [197, 195, 191, 174, 236, 226, 157, 196, 201, 196, 181, 164, 181, 229,␣
 ↪231]

t_statistic, p_value = stats.ttest_rel(before, after)

alpha = 0.05

print("T-statistic: ",t_statistic)
print("P-value: ",p_value)

if p_value < alpha:
    print("Reject the null hypothesis. The weight loss program is considered␣
 ↪effective.")
else:
    print("Fail to reject the null hypothesis. There is no significant evidence␣
 ↪that the weight loss program is effective.")
```

```
T-statistic:  6.6896995348736334
P-value:  1.0275656281485408e-05
Reject the null hypothesis. The weight loss program is considered effective.
```