Case Study 6 Homework: Exercises 1-4

Bookmark this page

Homework due Jul 14, 2021 05:59 +06

Exercise 1

1/1 point (graded)

In Exercise 1, we will calculate the chance homophily for an arbitrary characteristic. Homophily is the proportion of edges in the network whose constituent nodes share that characteristic.

How much homophily do we expect by chance? If characteristics are distributed completely randomly, the probability that two nodes x and y share characteristic a is the probability both nodes have characteristic a , which is the marginal probability of a squared. The total probability that nodes x and y share their characteristic is therefore the sum of the square of the marginal probabilities of each characteristic in the network.

Instructions

- Create a function `marginal_prob` that takes a dictionary `chars` with personal IDs as keys and characteristics as values; it should return a dictionary with characteristics as keys and their marginal probability (frequency of occurence of a characteristic divided by the sum of frequencies of each characteristic) as values.
- Create a function `chance_homophily(chars)` that takes a dictionary `chars` defined as above and computes the chance homophily (homophily due to chance alone) for that characteristic.
- A sample of three peoples' favorite colors is given in `favorite_colors`. Use your function to compute the chance homophily in this group, and store it as `color_homophily`.
- Print `color_homophily`.

Here's the code to get you started:

```
       from collections import Counter
import numpy as np

def marginal_prob(chars):
    # Enter code here!

def chance_homophily(chars):
    # Enter code here!

favorite_colors = {
    "ankit":  "red",
    "xiaoyu": "blue",
    "mary":   "blue"
}

color_homophily = chance_homophily(favorite_colors)
print(color_homophily)
```

What is the chance homophily of the `favorite_colors` dictionary?

0.555555555556 ✔

0.555555555556

Submit You have used 1 of 10 attempts

Hide Notes

Exercise 2

1/1 point (graded)

In the remaining exercises, we will calculate actual homophily in these village and compare the obtained values to those obtained by chance. In Exercise 2, we subset the data into individual villages and store them.

Instructions

- Note that `individual_characteristics.dta` contains several characteristics for each individual in the dataset such as age, religion, and caste. Use the `pandas` library to read in and store these characteristics as a dataframe called `df`.
- Store separate datasets for individuals belonging to Villages 1 and 2 as `df1` and `df2`, respectively.
- Note that some attributes may be missing for some individuals.
- Use the `head` method to display the first few entries of `df1`.

Here is the code to get you started:

```
      import pandas as pd

df  = pd.read_csv("https://courses.edx.org/asset-v1:HarvardX+PH526x+2T2019+type@asset+block@individual_cha
df1 = # Enter code here!
df2 = # Enter code here!

# Enter code here!
```

How many people had a `resp_gender` value of 1 in the first 5 entries of `df1`?

3 ✔

3

Submit You have used 1 of 3 attempts

Exercise 3

1/1 point (graded)

In Exercise 3, we define a few dictionaries that enable us to look up the sex, caste, and religion of members of each village by personal ID. For Villages 1 and 2, their personal IDs are stored as `pid`.

Instructions

- Define dictionaries with personal IDs as keys and a given covariate for that individual as values. Complete this for the sex, caste, and religion covariates, for Villages 1 and 2.
- For Village 1, store these dictionaries into variables named `sex1`, `caste1`, and `religion1`.
- For Village 2, store these dictionaries into variables named `sex2`, `caste2`, and `religion2`.

Here is some code to get you started:

```
sex1      = # Enter code here!
caste1    = # Enter code here!
religion1 = # Enter code here!

# Continue for df2 as well.
```

Hide Notes

What is the `caste` value for personal ID 202802 in village 2?
Enter your answer as letters only (no other characters).

OBC ✔

Submit You have used 1 of 10 attempts

✔ Correct (1/1 point)

Exercise 4

1/1 point (graded)
In Exercise 4, we will print the chance homophily of several characteristics of Villages 1 and 2.

Instructions
Use `chance_homophily` to compute the chance homophily for sex, caste, and religion In Villages 1 and 2. Consider whether the chance homophily for any attribute is very high for either village.

Which characteristic has the highest value of chance homophiliy?

- ○ Village 1, sex
- ○ Village 2, sex
- ○ Village 1, caste
- ○ Village 2, caste
- ○ Village 1, religion
- ◉ Village 2, religion ✔

Submit You have used 1 of 2 attempts

‹ Previous Next ›

edX for Business
Open edX
Careers
News

Legal

Terms of Service & Honor Code
Privacy Policy
Accessibility Policy
Trademark Policy
Sitemap

Connect

Blog
Contact Us
Help Center
Media Kit
Donate

Hide Notes