Forum Code Workbook

Assignment - 2 (ICS-214)

Problem 1. Raffle

(#Probability, #CompTools)

On a recent visit to Global Village in Dubai, three ZU students, Sara, Mahra, and Ahmad, attended a raffle event where there was a box with 1000 tickets, each with a unique number starting from 1 to 1000. A ticket was randomly taken out and the winning number was 1000. This led to the following discussion among Sara, Mahra, and Ahmad.

- Sara believes that it's very unlikely that such an easy number, 1000, was randomly drawn, and she suspects that the drawing must be rigged, hence a foul play.
- Mahra argues that while 1000 is the largest number but any number between 1 and 1000 is equally unlikely so no reason to suspect foul play.
- Ahmad, after listening to Mahra, got convinced and argued that there was a 0.1% chance of foul play.

Question 1 of 18

1a. Who do you think is right or wrong and why?

Normal B I U

Question 2 of 18

1b. Suppose if p represents the prior probability that the raffle is rigged then calculate the posterior probability of foul play.

Normal B I U

Question 3 of 18

1c. Assess Sara's strong disbelief in the raffle in the light of your answers to part 1b.

Normal

Question 4 of 18

1d. Assess Mahra's argument in the light of your answers to part 1b.

Normal

Question 5 of 18

1e. What do you get as answer of part 1b, when Ahmad's suggestion is taken?

Normal

Question 6 of 18

1f. In the light of above answers, present your findings in a way that all of three ZU students get convinced.

Normal

Code Cell 1 of 7

```
In [ ]
```

```
#OPTIONAL AS NEEDED
```

Run Code

Problem 2. Medical test

(#Probability, #MathTools)

Omer wants to be really certain about a diagnosis so he takes a series of identical medical tests. He hopes multiple tests will reduce his uncertainty.

Events:

D : Omer has the disease being tested for.
T_j : Omer tests positive on the j^{th} test for $j = 1, 2, ..., n$.

Let $p = P(D)$ be the prior probability that he has the disease.

Question 7 of 18

2.1 Assume for this part the test results are conditionally independent given Omer's disease status. Let $a_0 = P(T_j|D)$ and $b_0 = P(T_j|D^c)$, where a_0 and b_0 don't depend on j.

Find the posterior probability that Omer has the disease, given that he tests positive on all of the tests.

Hint: Since a_0 does not depend on j (the index of a test result), the algebra simplifies a lot since $P(T_i|D)P(T_j|D) = a_0^2$ for any test indices i and j.
The same holds for b_0.

Normal

Question 8 of 18

2.2 Suppose some people have a gene that makes them always test positive on this type of medical test. Let G be the event that Omer has the gene. Assume that $P(G) = \frac{1}{2}$ and that D and G are independent — that is, the gene does not make you more or less susceptible to the disease.

If Omer has the gene, he'll test positive on all tests.

If Omer does not have the gene, then the test results are conditionally independent given his disease status. Let $a_1 = P(T_j|D, G^c)$ and $b_1 = P(T_j|D^c, G^c)$, where a_1 and b_1 don't depend on j .

Now, suppose that Omer tests positive on all tests and find the posterior probability that Omer has the disease.

Normal

Question 9 of 18

2.3 Using the same setup as in part (2.), find the posterior probability that Omer **has the gene** given that he tests positive on all n of the tests.

Normal

Code Cell 2 of 7

```
In [ ]  #OPTIONAL AS NEEDED
```

Run Code

Code Cell 3 of 7

```
In [ ]  #OPTIONAL AS NEEDED
```

Run Code

Problem 3. Proofreading

(#Distributions, #ModelSelection, #MathTools)

A book has N typos. Two proofreaders, Sho and Haruna, independently read the book. Sho finds each typo independently with probability p_1, and Haruna with probability p_2. Let X_1 be the number of typos caught by Sho, and X_2 the number caught by Haruna, and let X be the number caught by at least one of the two proofreaders.

Question 10 of 18

3.1 Find the distribution of X. Motivate your answer.

Normal

Question 11 of 18

3.2 Assuming $p_1 = p_2$, find the conditional distribution of X_1 given that $X_1 = X_2 = t$. That is, if we know that t typos were found, how many were found by the first proofreader?

Hint: $X_1 + X_2 \neq X$. Note that we are conditioning on the number of typos found by Sho plus the number found by Haruna. This is not the same as the number of typos found by at least one of the two proofreaders — we are double-counting the typos found by both of them.

Normal

Question 12 of 18

3.3 You write a book with 100,000 words. On average, you make a typo once every 300 words. Sho and Haruna proofread your book. Sho finds 299 typos while Haruna finds 314 typos. Use model selection to decide whether Haruna is better than Sho at finding typos or whether it is just a coincidence that she found more than Sho this time.

Hints:

- Follow the frequentist model selection recipe.
- For the null model, think carefully about which distribution to use for the number of typos that exist in the book.
- You'll need your answer to part (2.) above.
- You will need to implement a simulation for this problem. Don't rely on looking up a known distribution of the test statistic.

Normal

Code Cell 4 of 7

```
In [ ]  #OPTIONAL AS NEEDED
```

Run Code

Code Cell 5 of 7

```
In [ ]  #OPTIONAL AS NEEDED
```

Run Code

Problem 4. Randomized response surveys

(#ParameterEstimation)

To ensure the privacy of a family participating in a survey, a ZU graduate uses the randomized response approach. He conducts a survey as follows by giving two types of envelopes:

- Envelope A: A proportion *p* of envelopes with "Is your family greater than 3?"
- Envelope B: Remaining (1-*p*) of envelopes with "Is your household income greater than 50000AED?"

Participants randomly choose to answer truthfully or untruthfully as follows:

- *t* is the probability they answer truthfully
- (1-*t*) is the probability they answer randomly (say "yes" or "no" with equal probability)

If *f* represents proportion of family greater than 3 and *i* represents proportion of family with income greater than 50000AED.

Question 13 of 18

4a. Find the Probability P(Yes) that a participant say "Yes" to the survey.

Normal

Question 14 of 18

4b. Determine the effect of truthfulness when all participants have income greater than 50000AED and family size bigger than 3.

Normal

Question 15 of 18

4c. What does the outcome reveal if all participants answer truthfully and the survey equally balances both questions?

Normal

Question 16 of 18

4d. If all participants answer absolutely randomly, does it eliminate the influence of the questions being asked?

Normal

Code Cell 6 of 7

In [ ]

```
#OPTIONAL AS NEEDED
```

Run Code

Code Cell 7 of 7

```
In [ ]    1  #OPTIONAL AS NEEDED
```

Run Code

Question 17 of 18

Problem 5: Reflection

#ProfessionalWorkProduct

Reflect back on one of your favorite questions in the assignment, explain how it expanded your Bayesian understanding, and identify the areas where you believe that you can use it in the future.

Normal

Question 18 of 18

OPTIONAL: if you prepared your solutions on another platform from which you can construct a PDF (for some or all of the problems) you can upload it here. For any problem/question whose solution is in the PDF please write "see PDF" in the appropriate workbook text entry box. And be sure that your answers are clearly labeled in terms of problem and part in the PDF.

Drop or upload a file here