ASSIGNMENT - 6

Vishanth Hari Raj B

Q.1:

PART A

	Ad	Probability
0	1	0.46882
1	3	0.39835
2	4	0.11646
3	2	0.01635
4	5	0.00002

Ad one has the highest probability.

PART B

	Ad	Mean	Std Dev	DOF	Probs
0	1	42.076923	1.391324	53	0.00000
1	2	57.289474	2.822110	39	0.00001
2	3	41.803922	1.945134	52	0.00000
3	4	47.888889	2.017655	46	0.00000
4	5	90.640000	6.137285	26	0.99999

Ad five has the highest probability.

PART C

	count	Ad
max_ind		
5	0.33845	5
2	0.20125	2
1	0.16850	1
4	0.15884	4
3	0.13296	3

Ad five has the highest probability.

QUESTION 2: (OPTIONAL)

Campaigns with lower CTR often tend to have higher post-click volumes because repeated exposures to the advertisement build recognition and trust among viewers. This familiarity increases the likelihood that users who eventually click on the ad are more engaged and interested, leading to higher conversion rates and greater post-click transaction volumes. Additionally, within the same category of campaigns (e.g., banner ads of the same size or video ads of the same duration), these ads resonate with a broader audience. While a smaller percentage of viewers may click on the ad initially, those who do click after multiple exposures are more likely to find the ad relevant and engaging, resulting in higher post-click volumes. This suggests that the ad effectively captures long-term interest and fosters deeper engagement, even if the immediate click rate is lower.

QUESTION 3: (OPTIONAL)

	Ad	Mean	Std Dev	DOF	count
0	1	42.076923	1.391324	53	0.16785
1	2	57.289474	2.822110	39	0.19760
2	3	41.803922	1.945134	52	0.13197
3	4	47.888889	2.017655	46	0.16018
4	5	90.640000	6.137285	26	0.34240

Using just these means, if we had used just the means themselves, we would have selected ad 1, which is different from the answer we got.

Consider the following example:

- **Campaign A**:
    - ObservedCTR: 0.1
    - MeanVolume: 100
    - Expected CTR*v: 10
- **Campaign B**:
    - ObservedCTR: 0.2
    - MeanVolume: 50
    - Expected CTR*v: 10

Although the expected values of CTR*v for both campaigns are the same (10), the observed CTR for Campaign B is twice as high as that for Campaign A. This suggests that Campaign B is more likely to have the highest CTR*v, even though their expected values are identical.

To directly answer the question, it is not always true that P(x>y) is greater than P(y>x)) when E(x) is greater than E(y). The two orderings will only be the same if the probability density functions for x and y are identical. In general, expected values do not fully capture the variability and distributional differences between random variables, which is why P(x>y)) can differ significantly from comparing E(x) and E(y).

Another Example:

Consider an extreme distribution x with two possible outcomes: 0 with probability 0.99 and 100,000,000 with probability 0.01. The expected value E(x) is very large due to the rare but extreme high value: E(x)=0×0.99+100,000,000×0.01=1,000,000

E(x) = 0 * 0.99 + 100,000,000 * 0.01 = 1,000,000

E(x)=0×0.99+100,000,000×0.01=1,000,000

Now, consider a regular distribution y between 0 and 1, with mean 0.5. The expected value E(y)=0.5

Although E(x) is much larger than E(y), y will be greater than x most of the time because x is 0 in 99% of the cases. Thus, P(y>x)≈0.99

This example demonstrates that expected values (means) can be misleading when skewed by extreme values. The probability P(x>y) considers the likelihood of outcomes rather than their magnitudes, offering a more accurate comparison. Consequently, direct multiplication of CTR and mean volume can yield different results compared to Bayesian probabilities, which account for the entire distribution of outcomes.

QUESTION 4 (OPTIONAL)

A) Prompts Entered into the Generative AI Tool

Prompt 1:

"I have two datasets: one with transaction profit volumes and one with clicks and exposures for ad campaigns. I need to calculate the Bayesian posterior probabilities for each campaign's true click-through-rate (CTR), average post-click transaction profit volume (v), and expected volume per impression (EVI). Can you guide me through the process using Python?"

Output: Steps to do it

Prompt 2:

"Can you help and assist in compute the following: The Bayesian posterior probability that the campaign's true click-through-rate is the highest across all campaigns, and the Bayesian posterior probability that the campaign's true average post-click volume per click is the highest across all campaigns. Use scipy.stats.beta.rvs and scipy.stats.t.rvs in Python."

Output: Python code pseudo code