Hypothesis Testing - Problems

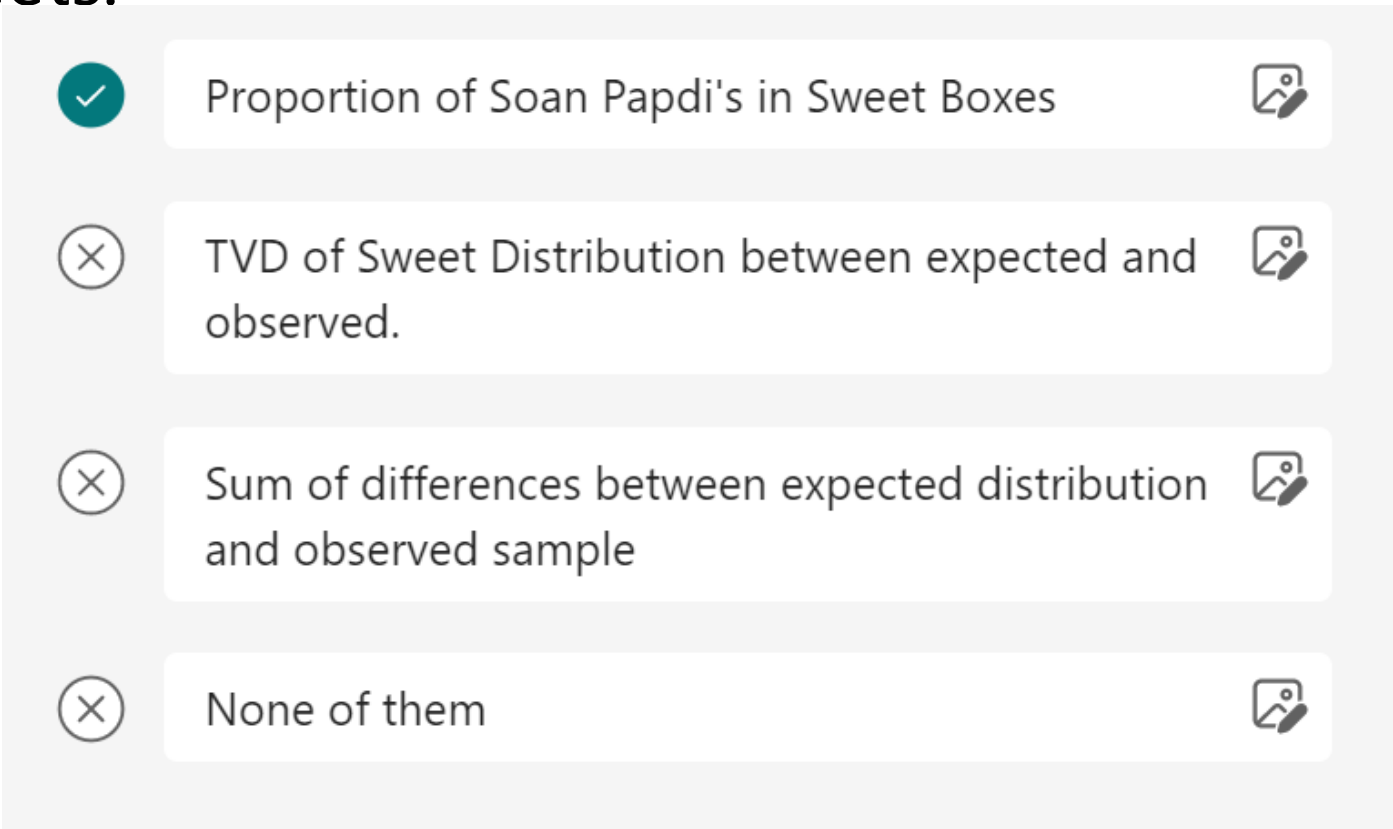
- XYZ Sweets Ltd has sweet making units in Chennai, Bengaluru, and Hyderabad. The unit located in Chennai produces Assorted Sweet boxes of varying numbers as per the following proportions

- Mysore Pak 0.6
  Motichoor Laddu 0.2
  Soan Papdi 0.2
- For Diwali, you recently received an Assorted Box with 10 sweets from your friend originally from Chennai. On opening, you discover 4 Soan Papdi's, much to your disappointment (as you are not so fond of it). However, you forgot to check the package cover before disposing of, as to whether it was from Chennai or elsewhere (also your friend travels frequently to Bangalore and Hyderabad), You decide to run a hypothesis test to figure out whether this Sweet Box came from Chennai unit or not?

  Select ALL the relevant statements implied by the NULL hypothesis you formulate based on the available information. You need to mark all the correct choices to be awarded marks for this question

- [x] Sweet box came from Chennai Unit.
- [x] Your sweet box is only a chance variation.
- [x] The chance of finding Soan Papdi's in the Sweet box is 20%
- [ ] The sweet box didn't come from Chennai unit
- [ ] The chance of knowing number of Soan Papdi's in the Sweet box is not known in advance
- [ ] The chance of knowing number of Soan Papdi's in the Sweet box is computable but not 20%

- For the above stated Sweet Box problem, identify a suitable test statistic that you can use for comparing standard sweet boxes with 10 sweets.



A scientist claims that in a particular region, 65% of bees are honey bees, 25% are bumblebees, and 10% are carpenter bees. You are suspicious of this claim, so you take a uniform random sample of 100 bees in the area. You get the following results from your random sample:

Bee Type	Number of bees
Honey bee	75
Bumblebee	15
Carpenter bee	10

- a) In order to determine whether the scientist's claim is actually true, you want to perform a hypothesis test. Which of the above statements could be a reasonable null hypothesis? Also identify the alternative hypothesis.

  *Null: Bees are distributed according to the probability distribution specied by the scientist. (*Each bee is chosen randomly, with a 60% chance of being a honey bee, 30% chance of bumblebee, 10% of carpenter bee)

  Alt: Bees are not distributed according to the probability distribution specied by the scientist.

- b) You decide to use the Total Variation Distance (TVD) between the empirical distribution (in the sample) and the probability distribution specified by the scientist as your test statistic. Suppose that the result of your computation after simulation runs is **empirical_pval = 0.0258**. Enumerate your decisions with respect to the null hypothesis at various significance levels, viz. 1% (highly significant level), 5% (significant level)
  - If we use a p-value cut-o of 1%, we should accept the null hypothesis.
  - If we use a p-value cut-o of 5%, we should reject the null hypothesis.