Q7.

We are given a discrepancy measure $D(P_1, P_2)$. We will now describe a test of significance using the above discrepancy measure.

1] We first state the null hypothesis H_0 : P_1 & P_2 are drawn from the same population.

2] We will now use the above discrepancy measure $D = D(P_1, P_2)$ where large values indicate evidence against the null hypothesis

3] Calculate the observed discrepancy $d_{obs} = D(P_1, P_2)$

4] Shuffle the sub-populations M times and calculate the observed p-value :

$$\text{p-value} = P_r(D \geq d_{obs} \mid H_0 \text{ is true}) \approx \frac{1}{M} \sum_{i=1}^{M} I(D(P_{1,i}^*, P_{2,i}^*) \geq d_{obs})$$

which basically means finding the probability that a randomly shuffled sub-population has a discrepancy measure (D) at least as large as what we observed (d_{obs}), given that H_0 is true. The smaller the p-value, the greater the evidence against the null hypothesis.