

1. You are a data scientist working for an e-commerce company. The marketing team wants to know whether offering a discount on products increases the average time users spend browsing the website. To investigate this, you randomly select 200 users and split them into two groups: one group sees discounted prices, and the other sees regular prices. After one week, you collect data on the average time spent on the website by users in both groups.

What type of statistical study is this, and why?

What is the research question being investigated?

2. Assume you have a simple binary classification model that classifies whether a given house classifies as a mansion or a regular house (not a mansion). Given below is the confusion matrix of the model's results. Based on the matrix below, calculate the accuracy, precision, recall, and F1 measures.

	MODEL PREDICTIONS		
ACTUAL RESULTS		Mansion	Not Mansion
	Mansion	40	30
	Not Mansion	25	35

Solutions:

Q1.

1. This is an experimental study because the researcher is actively manipulating the independent variable (whether users see discounted prices or regular prices) and controlling the study conditions to observe the effect on the dependent variable (average time spent on the website).
2. "Does offering a discount on products increase the average time users spend browsing the website?"

Q2:

a. $Accuracy = \frac{(TP+TN)}{N} = \frac{(TP+TN)}{TP+TN+FP+FN} = \frac{(40+35)}{(40+35+30+25)} = 0.5769$

b. $Precision = \frac{TP}{(TP+FP)} = \frac{40}{40+25} = \frac{40}{65} = 0.6154$

c. $Recall = \frac{TP}{(TP+FN)} = \frac{40}{40+30} = \frac{40}{70} = 0.5714$

d. $F1 = \frac{2 * Precision * Recall}{(Precision + Recall)} = \frac{2 * 0.6154 * 0.5714}{0.6154 + 0.5714} = 0.6088$