Exercise 1:

Part 1: Confusion Matrix and Accuracy

We are building models to detect a **rare disease** (**1 = sick (positive)** and **0 = healthy (negative)**). To analyze classification performance in detail, we use the confusion matrix. Below is an empty confusion matrix for a binary classification problem:

	Actual Positive	**Actual Negative**
Predicted Positive	(A)	(B)
Predicted Negative	(C)	(D)

(a) Assign the correct terms (TP, FP, FN, TN) to fields (A), (B), (C), (D) in the confusion matrix above and briefly explain what each term represents.

(b) Look at the row sums and column sums of the confusion matrix. What do these sums represent?

(c) Express the accuracy metric using TP, FP, FN, and TN.

(d) For each of the 8 metrics in the table below, categorize which question (Question 1 or Question 2) it answers, and write a one-line interpretation of what the metric measures. Many metrics naturally group into two categories based on what question they answer:

- **Question 1:** "Among all actual positives/negatives, how many did we correctly/incorrectly classify?" (Metrics condition on real class: denominators positives (P) or negatives (N))
- **Question 2:** "Among all predicted positives/negatives, how many were actually correct/incorrect?" (Metrics condition on predictions: denominators predicted positives (PP) or predicted negatives (PN))

Metric	**Formula**
Precision (Positive Predictive Value, PPV)	$\frac{TP}{TP+FP}$
False Positive Rate (FPR, Fallout)	$\frac{FP}{FP+TN}$
True Negative Rate (TNR, Specificity)	$\frac{TN}{TN+FP}$
False Omission Rate (FOR)	$\frac{FN}{TN+FN}$
True Positive Rate (TPR, Recall, Sensitivity)	$\frac{TP}{TP+FN}$
False Discovery Rate (FDR)	$\frac{FP}{TP+FP}$
Negative Predictive Value (NPV)	$\frac{TN}{TN+FN}$
False Negative Rate (FNR, Miss Rate)	$\frac{FN}{TP+FN}$

Part 2: Metrics Derived from the Confusion Matrix

Now we apply this to actual data. We have 10 patients with true labels and predictions from two models:

Patient ID	1	2	3	4	5	6	7	8	9	10
True label y	0	0	0	1	0	0	0	0	0	0
Model A	1	0	0	0	0	0	0	0	0	0
Model B	0	1	0	1	0	0	1	0	1	0

(a) For both models, compute the accuracy. Based only on accuracy, which model seems better?

(b) For both models, compute: TPR, TNR, Precision, and NPV. For each of these four metrics, identify which model performs better and discuss what this means in practice (in particular for which aspects of the confusion matrix each model performs well or poorly, and which model you would choose in this medical setting).

Part 3: F1 Score (Harmonic Mean of Precision and Recall)

Sometimes we want a single metric that balances precision and recall. We define the F1 score as the harmonic mean of precision and recall:

$$F1 = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}.$$

(a) Using your values for Precision and Recall from Part 2, compute the F1 score for both models. What result do you obtain for Model A, and how does the F1 formula behave in this case? Explain your answer.

(b) Why might we prefer the *harmonic* mean instead of the arithmetic mean for combining precision and recall?

Part 4: Thresholds and the Choice of 0.5

So far we have worked with hard predictions (0 or 1). But many models output probability scores instead. The choice of threshold (the cutoff above which we predict class 1) affects the confusion matrix.

Consider Model C with probabilistic outputs for 6 patients (see below). Compute and compare TPR and FPR at thresholds 0.5 and 0.3. Describe how lowering the threshold from 0.5 to 0.3 changes the confusion matrix and the derived metrics, and explain why the choice of threshold is important.

Patient ID	1	2	3	4	5	6
True y	0	1	0	0	1	0
Model C score	0.20	0.40	0.10	0.30	0.35	0.05