6. (3.5 points)

A logistic model was built to predict the probability of a claim being fraudulent. Consider the predicted probabilities for the 10 claims below to be a representative sample of the total model.

Claim Number	Actual Fraud Indicator	Predicted Probability of Fraud
1	Y	11%
2	N	23%
3	N	15%
4	N	70%
5	Y	91%
6	Y	30%
7	N	11%
8	Y	75%
9	N	58%
10	N	27%

a. (1 point)

Construct confusion matrices for discrimination thresholds of 0.50 and 0.25.

b. (1.5 points)

Plot the Receiver Operating Characteristic (ROC) curve with the discrimination thresholds of 0.50 and 0.25. Label each axis and the coordinates and discrimination threshold of each point on the curve.

c. (0.5 point)

Describe an advantage and a disadvantage of selecting a discrimination threshold of 0.25 instead of 0.50.

d. (0.5 point)

Describe whether a discrimination threshold of 0.25 or 0.50 is more appropriate for a line of business with low frequency and high severity.

CONTINUED ON NEXT PAGE

QUESTION 6	
TOTAL POINT VALUE: 3.5	**LEARNING OBJECTIVE(S): A4c,d**
SAMPLE ANSWERS	
Part a: 1 point	

Sample 1

Discrimination Threshold: 50%

Claim	Actual Fraud	Predicted Fraud	True Pos	False Pos	True Neg	False Neg
1	Y	N	0	0	0	1
2	N	N	0	0	1	0
3	N	N	0	0	1	0
4	N	Y	0	1	0	0
5	Y	Y	1	0	0	0
6	Y	N	0	0	0	1
7	N	N	0	0	1	0
8	Y	Y	1	0	0	0
9	N	Y	0	1	0	0
10	N	N	0	0	1	0
		Totals:	**2**	**2**	**4**	**2**

Discrimination Threshold: 25%

Claim	Actual Fraud	Predicted Fraud	True Pos	False Pos	True Neg	False Neg
1	Y	N	0	0	0	1
2	N	N	0	0	1	0
3	N	N	0	0	1	0
4	N	Y	0	1	0	0
5	Y	Y	1	0	0	0

6	Y	Y	1	0	0	0
7	N	N	0	0	1	0
8	Y	Y	1	0	0	0
9	N	Y	0	1	0	0
10	N	Y	0	1	0	0
		Totals:	**3**	**3**	**3**	**1**

Discrimination Threshold: 50%

	Predicted		
Actual	Fraud	No Fraud	Total
Fraud	2	2	4
No Fraud	2	4	6
			10

Discrimination Threshold: 25%

	Predicted		
Actual	Fraud	No Fraud	Total
Fraud	3	1	4
No Fraud	3	3	6
			10

Sample 2

Claim	Actual Fraud	Above .50?	Above .25?
1	Y	N	N
2	N	N	N
3	N	N	N
4	N	Y	Y
5	Y	Y	Y
6	Y	N	Y
7	N	N	N
8	Y	Y	Y
9	N	Y	Y
10	N	N	Y

Discrimination Threshold 0.50

True Positive	False Negative
2	2
False Positive	True Negative
2	4

Discrimination Threshold 0.25

True Positive	False Negative
3	1
False Positive	True Negative
3	3

Part b: 1.5 points

Sensitivity = True Positives / Total Positives

Specificity = True Negatives / Total Negatives

False Positive Rate = 1 - Specificity

Discrimination Threshold: 50%

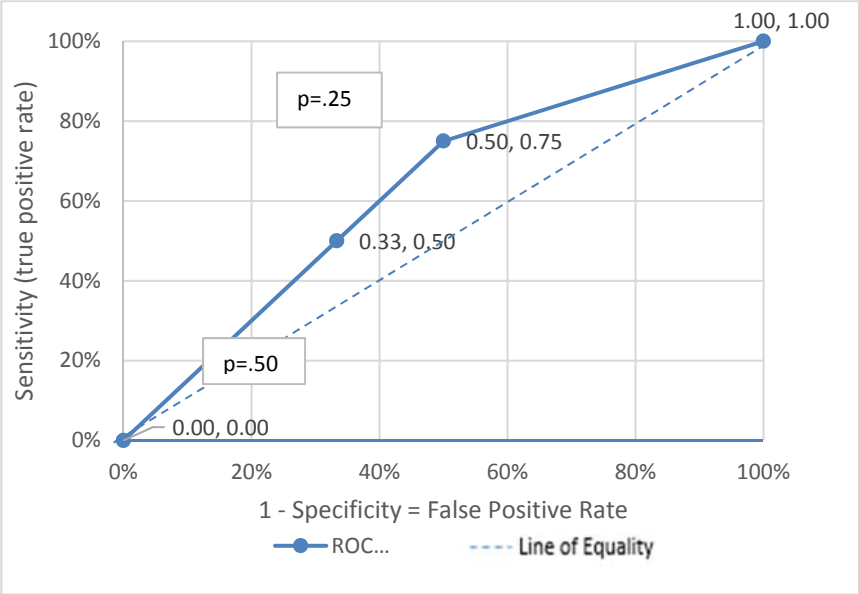
Sensitivity	0.500	=2/(2+2)
Specificity	0.667	=4/(2+4)
False Pos Rate	0.333	= 1 – 0.667

Discrimination Threshold: 25%

Sensitivity	0.750	=3/(3+1)
Specificity	0.500	=3/(3+3)
False Pos Rate	0.500	= 1 – 0.500

ROC Curve

false pos rate	sensitivity
X	**y**
0%	0%
33.33%	50.00%
50.00%	75.00%
100.00%	100.00%



Part c: 0.5 point

- An advantage of a lower threshold is that we will correctly identify more of the fraudulent claims. A disadvantage is that more of the non-fraudulent claims will be identified as fraudulent, so more time will be spent investigating.

- Advantage: You will catch more actual fraud claims because you will have a higher true positive rate. Disadvantage: You will have a higher false positive rate as well, which means you will waste resources to review claims that are not fraudulent.

Part d: 0.5 point

- A threshold of 0.25 is more appropriate. The high severity makes the cost of not investigating a fraudulent claim very high. The low frequency means that the number of additional claims that will need to be investigated is not very large. The cost of investigating these few additional claims is far less than the cost of potentially missing a few fraudulent claims at a higher discrimination threshold.
- Select 0.25 threshold. The benefit of detecting even one more fraud case is high (due to severity) and would outweigh the relatively low expense of inspecting a few extra false positives.

EXAMINER'S REPORT

Candidates were expected to demonstrate knowledge of confusion matrices, Receiver Operating Characteristic (ROC) curves, and the appropriateness of selecting various discrimination thresholds. This was a fairly straightforward question, and most candidates performed well. The question was very similar to the example from the reading.

Part a

Candidates were expected to construct two confusion matrices, one for each of the two discrimination thresholds. To receive full credit, candidates needed to show their work, such as creating a table of observations for true/false positive/negative occurrences, identifying the claims which fall into each matrix cell, or a sentence describing how the matrix was populated.

Common mistakes included:

- Not showing work and simply writing the correct values in the matrices

Part b

Candidates were expected to construct the ROC curve from (0,0) to (1,1) and passing through the points plotted at the two thresholds. Candidates did not have to plot every point along the curve; credit was given for piecewise linear and interpolated curved segments. The most common mistake was not including important components of the graph, such as axis labels, plotted points, or even the curve itself.

Part c

Candidates were expected to demonstrate knowledge of the advantages and disadvantages of selecting a particular discrimination threshold. To receive full credit, a sufficient level of detail was required. Simply stating that the 0.25 threshold increases sensitivity and decreases specificity, while true, does not adequately demonstrate the impact of the selection and only received partial credit.

Part d

Candidates were expected to discuss the reasoning behind selecting a particular threshold. To receive full credit, candidates needed to discuss the incentive for selecting a lower threshold, given the cost of missing a fraudulent claim due to high severity. The most common mistake was selecting the lower threshold without explaining why this made sense.