Ch03.08.A. CT Intracranial Mass Effect – Natural Language Processing

Since the brain is enclosed by the skull, swelling due to bleeding, infection, or inflammation is dangerous. This is called intracranial mass effect (IME). According to the Mortality Probability Model (Ch. 6, Box 6.2), IME is an important predictor of mortality in ICU patients.

IME is diagnosed by CT scan of the brain, but the results of the CT appear in the electronic health record (EHR) as a free-text report dictated by the radiologist; there is no <yes/maybe/no> field dedicated to IME. Enter Natural Language Processing (NLP), which may be used to extract <yes/maybe/no> results from free text in the EHR.

In the CTIME Study [REF], trained reviewers assigned <yes/no> labels for IME to 1557 head CT reports on ICU-bound emergency department patients. An NLP algorithm called "**term frequency–inverse document frequency"** (tf-idf) was trained on 1202 and tested on 355 reports. The algorithm assigns a score between 0 and 1 to each report. The authors created 3 score ranges: yes, > 0.5; maybe, 0.05 to 0.5; no, < 0.05. The test-set (validation) performance of the algorithm is summarized in the following table:

NLP Category	Score Range	Average Score	CT IME+ by trained reviewer		CT IME-- by trained reviewer		LR	NLP Yes/Maybe/No Category	
			N	%	N	%		Total N	%
Yes	>0.50	0.886	85	78.7%	8	3.3%	**24.20**	93	26.2%
Maybe	0.05 - 0.50	0.204	19	17.6%	44	17.9%		63	17.7%
No	<0.05	0.008	4	3.7%	195	78.9%	0.05	199	56.1%
	Total		108	100.0%	247	100.0%		355	100.0%

    a. In the test set of CT reports, what proportion were labelled by the (human) reviewers as IME+?

       108/355 = 30.4%

    b. What is the LR for the "Maybe" result category?

       17.6%/17.9% = 0.98 (~1.0 is fine)

    c. What proportion of reports categorized by the NLP algorithm as "Maybe" were IME+? We are asking for P(IME+|"Maybe").

       Shortcut, if the LR is ~1, you could just say ~30% -- your answer to (a)
       Or you can use the numbers. 19/63 = 30.2%

We are willing to accept the NLP categorization for IME when it is either "Yes" or "No", but we must manually review all reports labelled "Maybe".

d. What proportion of reports require manual review? We are asking for P("Maybe").
17.7%.

e. Does this depend (significantly) on the prevalence of IME+ in the sample?

No, it doesn't depend on the prevalence of IME+ in the sample (as long as the absolute number of IME+ is large enough that the discrete number of reports doesn't have an effect). This lack of dependence on prevalence is only the case for results with LR = 1. We can think of the overall proportion with a "maybe" result as a weighted average of those with that result who do and do not have IME, weighted by the prevalence of IME. But in this case, since the proportions are the same in IME+ and IME-, the weight (the prevalence of IME+) does not matter.

The table above is what we call an LR table. Here is the corresponding ROC table, partially completed.

	IME+	Sensitivity	IME-	1 - Specificity
>0.50			8	3.2%
>=0.05			52	21.1%
>=0	108	100.0%	247	100.0%

f. Finish completing the ROC table.

Begin Answer

	IME+	Sensitivity	IME-	1 - Specificity	
>0.50	*85*	*78.7%*	8	3.2%	0.0127
>=0.05	*104*	*96.3%*	52	21.1%	0.1559
>=0	108	100.0%	247	100.0%	0.7749

End Answer

g. How many line segments in the resulting ROC curve?

There are 3 segments in the resulting ROC curve.

h. (Extra Credit) What is the AUROC corresponding to the table in (e)?

Begin Answer:

	IME+	Sensitivity	IME-	1 - Specificity	(Sp1-Sp0) × (Se1+Se0)/2
	0	0.0%	0	0.0%	
>0.50	85	78.7%	8	3.2%	0.0127
>=0.05	104	96.3%	52	21.1%	0.1559
>=0	108	100.0%	247	100.0%	0.7749
					0.9435

AUROC = 0.9435

To do this with Stata, you first need to put the data in a frequency table, like we did for Kappa:

Score	IME	Freq
2	1	85
2	0	8
1	1	19
1	0	44
0	1	4
0	0	195

Then you can run:
```
. tabu score ime [weight=freq]
(frequency weights assumed)

           |        IME
   Score |      0          1 |     Total
-----------+----------------------+----------
       0 |      195          4 |       199
       1 |       44         19 |        63
       2 |        8         85 |        93
-----------+----------------------+----------
   Total |      247        108 |       355

. roctab ime score [weight=freq]
(frequency weights assumed)

                     ROC                    -Asymptotic Normal--
        Obs         Area      Std. Err.    [95% Conf. Interval]
       --------------------------------------------------------------
        355        0.9435       0.0131       0.91784      0.96910
```

End Answer