BIOS 6611 Homework 5

Due Tuesday, October 8, 2019 by 11:59 pm to Canvas Assignment Basket

Evaluating Diagnostic Tests*

A new instrument—the Chinese Mini-Mental Status Test (CMMS)—was proposed to identify dementia among people in China. Participants completed the CMMS, alongside an extensive clinical evaluation by psychiatrists and nurses to make a definitive diagnosis of dementia. The table below shows the results for one subgroup of participants (participants with some formal education, who were generally likely to score higher on the CMMS than participants without formal education).

	Clinical Diagnosis	
CMMS Score	**Non-Demented**	**Demented**
0-5	0	2
6-10	0	1
11-15	3	4
16-20	9	5
21-25	16	3
26-30	18	1
Total	46	16

A) Use R to calculate the sensitivity and specificity of the test when a CMMS score of ≤20 is used to identify people with dementia.

B) The cut-off value of ≤20 is arbitrary. Make a table showing sensitivity and specificity for cut-off values of 5, 10, 15, 20, 25, or 30. Use R to obtain the values. You can use whatever software you like to make the table, but simply printing out a matrix or data frame in R with all the values in it is fine, too.

C) In the context of this specific diagnostic test—a survey instrument to identify dementia—brainstorm a possible consequence of a false positive, then a possible consequence of a false negative.

* Problem adapted from Rosner's *Fundamentals of Biostatistics*. Data from Katzman, R., Zhang, M., Wang, Z., Liu, W.T., Yu, E., Wong, S.C., Salmon, D.P. and Grant, I., 1988. A Chinese version of the Mini-Mental State Examination; impact of illiteracy in a Shanghai dementia survey. *Journal of clinical epidemiology*, *41*(10), pp.971-978.

D) Based on your table in part B, select a cut-off for the CMMS, assuming CMMS false positives and false negatives are equally undesirable (in reality, the errors might not be equally problematic, as we saw in Part C).

E) Plot the ROC curve and obtain the AUC using R. How well do you think the CMMS discriminates between people with and without dementia?

Use data manipulation functions to create a data frame (e.g. data.frame() along with the rep(), cbind() and rbind() functions) such that each row represents two variables for each person in the dataset: the score on the CMMS and whether or not a person has dementia. Apply the Epi package in R to plot the ROC curve and obtain the AUC (you can refer to the R code in Lecture 9).

F) Extra Credit: Plot the ROC curve using your answers in B. Use programming statements to implement the trapezoidal rule to obtain the AUC. How similar is it to the AUC from part E?

G) For a cutoff of ≤20, what would the positive and negative predictive values (PPV, NPV) of the CMMS be in a Chinese population with dementia prevalence of 10%? Of 40%? Comment on the difference in the sets of predictive values for the two prevalence values.

H) Extra Credit: In contrast to the PPV and NPV, likelihood ratios (LR+ and LR-) are measures of a test's clinical utility that do not depend on prevalence. (Note: A rule of thumb is "LR+ ≥ 5" helps "rule in" disease, whereas LR- ≤ 0.2 helps "rule out" disease.)

i. Calculate the LR+ and LR- for each of the cut-offs in Part B. Summarize what you observe. Is there a single cutoff that makes the CMMS a good test for both ruling in and ruling out dementia?

ii. Assuming prior odds of dementia of 0.3, obtain the posterior odds of dementia and the posterior odds of no dementia for the various combinations of LR+ and LR-. Summarize what you observe.