this spreadsheet works just the same way as the example shown in the *Bombers and Seagulls*Spreadsheet (and the video) but has 10,000 rows of data instead of 20. It is designed to provide a realistic simulation of the cost-benefit assumptions that must be made to set the classification threshold for a medical diagnostic product. You can download the spreadsheet by clicking on, "Cancer Diagnosis.xlsx."

[Cancer Diagnosis.xlsx](https://d3c33hcgiwev3.cloudfront.net/_80a72b1369dfd15319274bec0f5f5f92_Cancer-Diagnosis.xlsx?Expires=1489708800&Signature=Klpz6HSG-57P2fLvWaFW0Ea6bRVmEAL6TMcsbqCH25L47vHFNorJ3LIkpMWbmvfu5TUIY1bvB-rxnJ0SooyBEF5-SbfE7SU9zc6GjdFuYIYhgIYxTRbocVDvBuCTaPZW5JVALCNzrdHXGJPOkHeohRKo8zxzquNbH7Ws0n2Q0w4_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

The *ranked scores* in Column A are the level of a certain protein as measured by the diagnostic test.

The true Condition for each protein level score is given in Column C [Cancer =1, No Cancer = 0].

A threshold for positive classification can be set between any two protein levels. For each threshold given in Column A, the resulting number of False Negative classification errors is given in the same row of Column H, and the False Positive classification errors in the same row of Column F.

This spreadsheet is designed to allow you to observe how changing cost inputs impacts both (a) the overall costs of using a cancer diagnostic test at each threshold, and (2) what threshold should be chosen as optimal (minimizing cost).

The cells to input the assumed costs of classification errors are Cell G3 for cost per False Negative (missing a case of cancer) and H3 for cost per False Positive (a false alarm). Total costs at each threshold are given in Column K.

The minimum total cost, and minimum cost per event (per diagnostic classification reported), are displayed in cells K4 and L4, and the optimum threshold – the lowest protein level score that should be classified positive - is displayed in cell M4.

At the default costs of $50,000 per False Negative error and $500 per False Positive error, the minimum cost per event is $119.90, and the optimal threshold for positive classification is 16551.930.

Try changing the inputs yourself. If you keep the cost per False Negative the same, but raise the cost per False Positive, would you in general expect the new optimum threshold to be higher (fewer total positive classifications) or lower (more total positive classifications)?

Change the cost per False Positive to $1500 and you will see that the optimum threshold is *higher.* The threshold for the first Positive classification moves from 16551.930 [item ranked 2094] to 16824.137 [item ranked 1822].

Similarly, if you reset the cost per FP at $500 and raise the cost per FN to from $50,000 to $500,000, the optimal threshold is lower – it falls to 13307.537 [item ranked 5338]

Questions:

Use the Cancer DiagnosisSpreadsheet to answer Questions 3 and 4.

[Cancer Diagnosis.xlsx](https://d3c33hcgiwev3.cloudfront.net/_80a72b1369dfd15319274bec0f5f5f92_Cancer-Diagnosis.xlsx?Expires=1489708800&Signature=Klpz6HSG-57P2fLvWaFW0Ea6bRVmEAL6TMcsbqCH25L47vHFNorJ3LIkpMWbmvfu5TUIY1bvB-rxnJ0SooyBEF5-SbfE7SU9zc6GjdFuYIYhgIYxTRbocVDvBuCTaPZW5JVALCNzrdHXGJPOkHeohRKo8zxzquNbH7Ws0n2Q0w4_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

This spreadsheet gives 10,000 pairs of scores – the level of a (fictional) cancer diagnostic protein in Column A - along with the actual condition: 1 = cancer, 0 = no cancer in Column C.

Change the cost per False Negative classification to $20,000 [cell G3]. Change the cost per False Positive classification to $1,000 [cell H3].

**Question: What is the new minimum cost per event/cost per test (rounded to the nearest dollar)?**



$181



$187



$183



$185

1  
point

4.

What is the lowest level of protein that should be classified "Positive" to achieve the minimum cost per test at the new costs per error given above?



18213.7



18202.407



18204.545



18204.498

Can a change in classification threshold change a diagnostic test's True Positive Rate? Use logic - no need to calculate any numbers.



No



Yes

1  
point

6.

"Condition Incidence" is the portion of a population that actually has the Condition being studied. Can a change in threshold change the Condition incidence? Use logic - no need to calculate any numbers.



No



Yes

1  
point

7.

Does the change in threshold change the test’s “classification incidence” (also called “test incidence”)? Use logic - no need to calculate any numbers.



No



Yes

1  
point

8.

Does the change in threshold change the test's Area under the ROC Curve? Use logic - no need to calculate any numbers.



Yes



No

1  
point

9.

Use the Forecasting Soldier Performance Spreadsheet to answer this question:

[Forecasting Soldier Performance.xlsx](https://d3c33hcgiwev3.cloudfront.net/_18267f8cf82cde5920753ca6efa3589e_Forecasting-Soldier-Performance.xlsx?Expires=1489708800&Signature=YfyzL~ueM6iUayqpX-TXD4SAYGetduIGw8PwFx6mhMy47oRinZbvMqHmj0VdzCktTWn86-UdL0~9AnWGC3Z3zsYNMzFYOd-clCnMVEWYQMT0eNBCzKluak5LugmweDeQJ6WNm0ot-c3sbfxknBYDBJ5GYBvmrY4-d3T9tOHFAtU_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

What is the False Positive Rate if we use the sum of standardized height and standardized weight as the score; and set a threshold at -1.28?



0.67



0.4



0.6



0.33