**HOW TO CONSTRUCT ROC CURVES**

by Simon Moss

|  |
| --- |
| **Introduction** |

Researchers often utilize instruments or procedures that are designed to predict which of two, or sometimes more, outcomes are likely. For example,

* researchers might evaluate an instrument that is designed to diagnose some illness or problem, such as diabetes or dementia
* researchers might develop a procedure to predict which research papers are likely to be accepted or rejected.

A technique called ROC curves tend to be applicable in these circumstances. To illustrate, in practice, Alzheimer’s Disease is hard to diagnose. To diagnose Alzheimer’s Disease definitively, practitioners need to complete expensive physiological tests. So, some researchers want to develop a simple measure—perhaps a series of 20 questions—to diagnose this disease. So, these researchers

* administer this test to 200 elderly participants
* subject these participants to expensive tests to assess whether or not they should be diagnosed with Alzheimer’s Disease
* the following table presents an extract of the results

These results imply the measure could be effective in the future. If you scan the columns you will notice an obvious pattern: the individuals who score high on this measure are also more likely to be diagnosed with Alzheimer’s Disease. Conversely, the individuals who score low on this measure are not as likely to be diagnosed with Alzheimer’s Disease. Nevertheless, a casual scan of these data is not enough to answer two important questions:

* First, how accurate is this measure? Is this measure more or less accurate than alternative instruments or procedures that are similar in cost?
* Second, what criterion should be used to predict Alzheimer’s Disease. Should practitioners conclude that scores over 8, 10, 12, or what predicts Alzheimer’s Disease?

The approach called ROC curves, or receive operating characteristic curves, can answer the first question. These ROC curves can also uncover insights that help answer the second question as well.

|  |
| --- |
| **Output of this technique** |

 Before we discuss how to conduct this technique as well as the rationale that underpins this technique, we will first outline how to interpret the output. In particular, this technique will generate a graph and an index called the *area under the curve*. Here are four examples of this graph together with the areas under each curve.

**What are these areas under the curve?**

 The area under each curve represent the proportion of area of this square that is under—or, more precisely, to the right of these curves. In the first graph, 0.9 of the area of this square is under or to the right of this curve. In the four graph, 0.6 of the area of this square is under or to the right of this curve.

**How should you utilize these areas?**

 Simply, high numbers indicate the measure, instrument, or procedure accurately diagnoses or predicts the outcome. In particular, 0.9 represents an excellent measure; 0.8 represents a good measure; 0.7 represents a fair measure; and 0.6 or under represents an inadequate measure. Therefore, in this instance, the researchers hope their measure of Alzheimer’s Disease generates a graph that represents the first figure instead of the last figure

**Which criterion should we utilize?**

 The graph also offers some insight into the criterion or cut-off. That is, in this example, the graph could help researchers decide which score should differentiate people who should be diagnosed with Alzheimer’s Disease from people who should not be diagnosed with Alzheimer’s Disease. Specifically, researchers could

* locate segments of graph that are relatively flat or horizontal
* the beginning of these segments are often suitable criteria
* however, in practice, more important considerations, such as the implications of overlooking a disease or falsely diagnosing a disease guide these decisions.

**Other data**

 Most statistical packages that generate ROC curves, such as SPSS, also display other output. For example, these packages will tend to present a table that specifies the coordinates of this graph.

|  |
| --- |
| **How to generate ROC curves** |

 ROC curves are, generally, easy to generate. This section demonstrates how to generate ROC curves in SPSS. But, the procedure is equally simple in most statistical packages. In particular, in SPSS, choosing *Analyze* and then *ROC curve*—located towards the bottom of this list—will generate the following screen.

* In the box called *Test Variable*, specify the column that represents the score from the specific measure, instrument, or procedure
* In the box called *State Variable*, specify the column that differentiates the two outcomes.
* In general, next to *Value of State Variable*, enter the number 1.
* That is, whenever possible, construct the data so that 1 represents the outcome you are attempting to predict or diagnose, such as Alzheimer’s Disease or acceptance of papers. 0 represents the other alternative, such as no Alzheimer’s Disease or non-acceptance of papers
* Tick all the other boxes on this screen.
* Press *Options* to generate the following screen
* To decide whether these options are correct, determine whether higher scores on your measure, instrument, or procedure indicate the outcome you are attempting to predict or diagnose.
* In this example, higher scores are more likely than lower scores to indicate Alzheimer’s Disease.
* Otherwise, you would need to shift the option from *Larger test results indicates more positive test* to *Smaller test results indicates more positive test*
* Note that, because you are attempting to diagnose Alzheimer’s Disease, this diagnosis is referred to as the positive classification. But obviously, wwe are not implying that Alzheimer’s Disease is a positive experience.

|  |
| --- |
| **The underlying rationale** |

 But, what is the rationale that underpins these ROC curves? How do statistical techniques generate these curves? To appreciate this rationale, consider the following pair of graphs. The first graph represents the number of participants with Alzheimer’s Disease that generated a score of 1, 2, 3, and so forth on this measure. The second graph represents the number of participants without Alzheimer’s Disease that generated a score of 1, 2, 3, and so forth on this measure. These graphs are called histograms or frequency distributions.

 The next figure merely blends the previous graphs into one graph. Furthermore, we have inserted a vertical line on this graph. The relevance of this vertical line will be clarified soon.

So, how can we convert this graph into an ROC curve? To illustrate, suppose we decided this line represents the criterion or cut-off. That is, participants who generated scores about this line are diagnosed with Alzheimer’s Disease. Participants who generated scores below this line are not diagnosed with Alzheimer’s Disease. We can now compute two indices, called sensitivity and specificity. In essence

* **sensitivity** represents the extent to which the test can identify people with the diagnosis or outcome, such as Alzheimer’s Disease or accepted papers
* in this example, sensitivity is the proportion of participants with Alzheimer’s Disease whose scores exceed the criterion
* **specificity** represents the extent to which the test correctly rejects people who should not be diagnosed or conferred the outcome
* in this example, specificity is the proportion of participants without Alzheimer’s Disease whose scores are lower than this criterion

To illustrate, in this example

* of the people with Alzheimer’s Disease, about 60% exceed the criterion and would be diagnosed; so sensitivity would be 0.6
* of the people without Alzheimer’s Disease, about 80% are less than the criterion and would not be diagnosed; so specificity would be 0.8
* can identify participants in the positive classification; in this instance, sensitivity
* proportion of participants with Alzheimer’s Disease whose
* first, we need to compute a measure called sensitivity; that is, what proportion of individuals with Alzheimer’s Disease in our sample would be diagnosed with Alzheimers

|  |  |  |
| --- | --- | --- |
| Main approaches | Key features of this approach | When is this approach applicable |
| Random effects models—including random intercepts and random slopes | * Assumes that each irrelevant cluster—such as each school— differs randomly from each other on the measures
 | * When you have recruited your sample of participants, animals, specimens, or whatever from several broader clusters
* When you measure the same people, animals, or units more than once over time
 |
| Correlations across repeated measures | * Considers the correlations across time
 | * When you measure the same people, animals, or units more than once over time
 |