Confusion Matrix DEI:
Using Implicit Bias to Master Statistical Classification

Robert J. Padgett

The Problem

The Confusion Matrix

The aptly named confusion matrix is the primary tool used to summarize the results of classification procedures widely used in the health, behavioral, and data sciences. In the health sciences, the tool is used to evaluate the accuracy and efficiency of a diagnostic test, in psychology it is used to summarize the results of a signal detection task, and the method is the primary tool for assessing prediction outcomes of a machine learning algorithm in the data sciences.

		Actual Class
Predicted Class	No	Yes
No	True Negative (TN)	False Negative (FN)
Yes	False Positive (FP)	True Positive (TP)

Note:

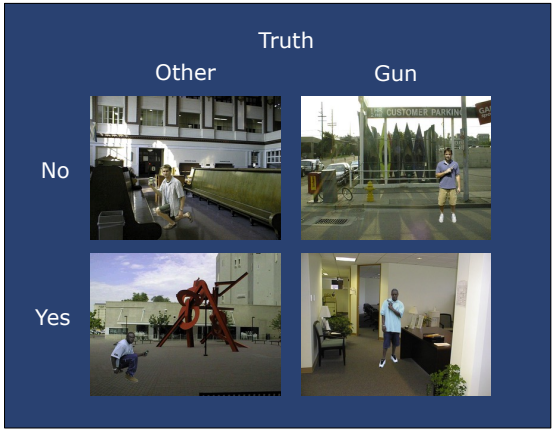
The words "Negative" and "Positive" refer to the type of prediction (Yes or No) The words "True" and "False" refer to the accuracy of that prediction. Students have trouble understanding of the difference between Sensitivity and Specificity which is *obsessively* the focus of this exercise. The many additional measures that can be calculate from a confusion matrix just compounds the problem for student understanding.

The Measures:

- Accuracy = (TN+TP) / N
- No Information Rate (FN+TP) / N (or the opposite)
- Sensitivity = TP / (TP+FN)
- Specificity = TN / (TN+ FP)
- Precision (PPV) = TP / (TP+FP)
- NPV = TN / (TN+FN)
- AUC – Sensitivity relative to Specificity

The Task

Do you see a gun?



Student Report

Confusion Matrix

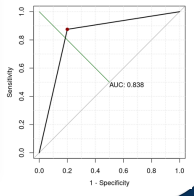
	Other Object (Noise Only)	Object was a Gun (Noise + Signal)
"No, I saw someting else"	True Negative (CR) 32	False Negative (Miss) 5
"Yes, I saw a gun"	False Positive (FA) 8	True Positive (CD) 35
Total	40	40

Complete Confusion Matrix Results

```
## Confusion Matrix and Statistics
##
##           reference
## prediction Object Gun
##     Object     32   5
##     Gun         8  35
##
##                Accuracy : 0.8375
##                  95% CI : (0.7382, 0.9105)
##     No Information Rate : 0.5
##     P-Value [Acc > NIR] : 0.0000000003208
##
##                   Kappa : 0.675
##
##  Mcnemar's Test P-Value : 0.5791
##
##             Sensitivity : 0.8750
##             Specificity : 0.8000
##          Pos Pred Value : 0.8140
##          Neg Pred Value : 0.8649
##              Prevalence : 0.5000
##          Detection Rate : 0.4375
##    Detection Prevalence : 0.5375
##       Balanced Accuracy : 0.8375
##
##        'Positive' Class : Gun
```

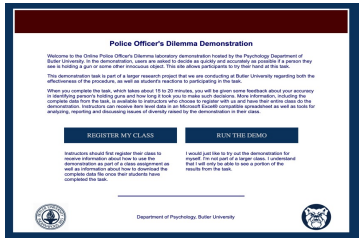
In-Class Canvas Assignment

Individual ROC Curve



How can YOU use it?

https://pod.butler.edu



	Accuracy	Sensitivity*	Specificity*
White	0.84	0.83	0.85
Black	0.85	0.90	.80

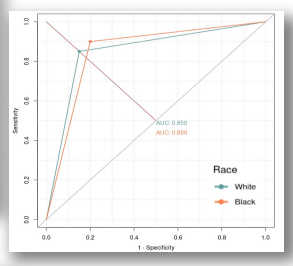
Note:

To receive anonymous student results by stimulus race you must register your class for the task and download the data. The site includes templates for data analysis using Excel, SPPS, and R Studio. An instructor guide for how to discuss implicit bias in the classroom is available at the site.

Past research demonstrates this task can be useful for both teaching about implicit bias and understanding confusion matrix results. Results typically show no difference by race on overall accuracy. However, higher levels of sensitivity (i.e., accuracy on "Gun" trials) and lower levels of specificity (i.e., accuracy on "Other Object" trials) when the depicted person is black are common.

Reaction time results (provided with the task data) also support the role implicit bias play in participant performance. Participants are faster at identifying objects that are guns overall, but especially when the stimulus person is Black. It takes longer to identify other objects when held by black individuals.

Class ROC Curves



Class RT Data