

**Outperforming a Multi-Level Classification Model
Whose Chance Performance is Large
Bruce Ratner, Ph.D.**

Situation

A classification model is to be built (sample size is 100,000) for assigning individuals into one of three classes, **a**, **b**, and **c** whose actual percents are 17%, 12%, and 71%, respectively.

Before starting out to build, it is good to know what one is up against. I peek at the chance model corresponding to the classification model of the noted actual percents. The classification matrix of the chance model is in Table 1, below. By chance, just randomly assigning the 100,000 individuals into the nine cells of Table 1, one expects to get a *total correctly classified percent* (TCC_PCT) of 54.74% $(=(2890+1440+50410)/100000)$.

Table 1.
Classification Matrix of the Chance Model
of Actual Percents 17%, 12%, and 71%

actual		predicted			
Frequency	Percent				
Row Pct	Col Pct	a	b	c	Total
a		2890			17000 17.00
b			1440		12000 12.00
c				50410	71000 71.00
Total		17000 17.00	12000 12.00	71000 71.00	100000 100.00

Issue At-Hand

Should the statistician embark on a modeling exercise when a chance model is large? With a large TTC_PCT, how much improvement over chance can realistically be expected? If a maximum classification (MC_PCT) is sought, why not assign all the individuals to class **c**? In this case, MC_PCT=71%

This [article](#) illustrates a strategy of build to significantly beat a “big” chance model of a multi-level classification model.