

Novometric vs. Logit vs. Probit Analysis: Using Gender and Race to Predict if Adolescents Ever Had Sexual Intercourse

Paul R. Yarnold, Ph.D.

Optimal Data Analysis, LLC

Prior research modeled if adolescents ever had sexual intercourse (the dependent variable: yes=1, no=0) using main-effect logit and probit analysis models treating gender (female=0, male=1) and race (“white”=1; “black”=2) as the independent variables. Both models identified statistically significant effects for gender and race, and both models misclassified all observations that had sexual intercourse: ESS=0. In exploratory novometric analysis conducted for these data a statistically significant, cross-generalizable race effect emerged that yielded moderate ESS=25.25, $p<0.001$.

SAS™ code used to construct the data analyzed herein¹ is given in the Appendix. Novometric, ODA, and EO-CTA analyses² identified a single two-strata model that had stable classification performance in LOO analysis: if race=“white” predict sexual intercourse=no; otherwise predict sexual intercourse=yes. Table 1 presents the confusion matrix for this model.

Table 1: Novometric Model Confusion Matrix

		Predicted Intercourse		
		No	Yes	
Actual Intercourse	No	283	59	82.8%
	Yes	69	51	42.5%

As seen, the model accurately classified 7 in 8 adolescents who did not have intercourse, but only 2 in 5 adolescents who did have intercourse (50% sensitivity is expected by chance

for each class category in two-category designs without analytic weights²). These findings show that race is potent for predicting adolescents who abstain, but impotent for predicting adolescents who have intercourse.

References

¹Liao TF (1994). *Interpreting probability models: Logit, Probit, and other generalized linear models*. Beverly Hills, CA: Sage (pp. 12-25).

²Yarnold PR, Soltysik RC (2016). *Maximizing predictive accuracy*. Chicago, IL: ODA Books. DOI: 10.13140/RG.2.1.1368.3286

Author Notes

This study analyzed publically available data. No conflict of interest was reported.

Appendix

SAS™ Code used to Construct (Reproduce¹) the Data File for Analysis by ODA Software²

```
data real;                                Do n=1 to 134;                            Do n=1 to 23;
infile datalines;                          put '1 1 0';                              put '2 1 0';
input group row                             end;                                        end;
column;                                     Do n=1 to 26;                              Do n=1 to 22;
cards;                                     put '1 0 1';                              put '2 0 1';
1 1 1                                       end;                                        end;
;                                           Do n=1 to 149;                            Do n=1 to 36;
Data example;                              put '1 0 0';                              put '2 0 0';
Do n=1 to 43;                              end;                                        end;
put '1 1 1';                               Do n=1 to 29;                              Output;
end;                                       put '2 1 1';                              Run;
end;                                       end;
```