**Classifier Variance Analysis by Self-Report**

This 2-pager summarily describes what was done, why it was done, and what I found. What was found takes the form of 5+ talking points.

**I. What Was Done**

A survey was conducted in which respondents utilized the NamePrism service[[1]](#footnote-1) and reported whether they agreed with the ethnicity estimate produced by NamePrism. Self-reported agreement was correlated to a number of other factors assembled in the survey.

80 responses were obtained between May 14 and 15, 2018.

**II. Why It Was Done**

Previous research[[2]](#footnote-2) attempted to directly assess accuracy of ethnic classifiers. That research provided low confidence findings attributable to small sample size. The present paper is able to provide a much larger sample because of a change in the questioning strategy.

SurveyMonkey prohibits asking for personally identifiable information when using a purchased audience. This prohibited the previous paper from utilizing a purchased audience because the previous paper required a LinkedIn profile URL. The current paper, in contrast, asks the respondent to utilize the name-prism service themselves and report the result. This introduces methodological benefits and risks, and it is not a strictly preferred methodology in theory. This approach is considered independent, and results from both studies is synthesized to provide results which combine gains from both approaches with cross-mitigated risks.

**III. What I Found**

1. Sample was 85% white and about 50% male. This is more balanced than the previous study but still largely white. The sample is all US residents.
2. About 50% of people agree with the ethnicity given by NamePrism
   1. This contrasts noticeably with the LinkedIn-based study. Some of this may be attributable to a “Salah effect,” wherein disagreement is not with the best-selection among available options, but rather, with the list of available options. Salah was a Middle Eastern man who chose White among the given options, and this agreed with the classifier.
3. In this sample, only about 50% of people stated they believe their last name
4. Under linear regression, gender does not have a significant effect. Some regions had significant effects. While ethnicity didn’t have a significant effect, the most significant effect was a negative effect on black ethnicity. Black ethnicity also had the largest ethnic effect size.
   1. The hypothesis that majority ethnicity would be best-predicted fails, but without confidence. Native Americans had the most accurate ethnicity predictions, but there was only 1 respondent who identified as Native American. Whites had the second smallest ethnic error size. It may still be that majority ethnicities are generally well-predicted, particularly with common names in society, while some minorities outperform due to unmistakably strong name-ethnicity matching and/or a reluctance to utilize common names.
   2. That is, a finding: Nominal minorities and extreme minority groups may be importantly different. Minorities may be heterogeneously integrated into society such that nominal minorities adopt common names, leading to name-base classification error, while extreme minorities reject common names, leading to effective name-based classification.
   3. Under linear regression, \_const > 1, which is logically impossible or at best non-intuitive.
5. Probit had a slightly higher pseudo r2 compared to logit.
   1. Under probit, regional effects were attenuated. 4 regions few had p < .36.
   2. The Native American respondent was dropped, leading to White ethnicity being best-predicted. The negative black effect became slightly more significant.
   3. Linear age had a weak effect (p=.45). Marginal and cubic effects were also tested. When linear and squared age are both included on the right hand, both factors obtain p < .2, a noticeable finding considering the sample size. Inclusion of the cubic removes any important relation.
   4. Pseudo-r reached .1085 in the preferred model
   5. In the preferred model, the constant has a value of -1.6. A simplified interpretation of this constant from an odds-ratio perspective would be to say that the model suggests a baseline case respondent is less likely agree than to disagree with the classifier ethnicity estimate. A baseline case is a respondent who obtains a value of 0 for each factor. That is, a person who is not male, age 0, from an omitted region, and of an omitted ethnicity, such as a white person.
   6. M3 is a linear implementation of the preferred probit model, used only to facilitate interpretation. It suggests a model constant value of .35, which is interpreted as about 35% rate of agreement in the baseline case.
6. In conclusion, a simple `sum voi` suggests self-reported ethnicity agrees with classifier estimates about 50% of the time, while regression analysis indicates that certain factors explain away some of this agreement, leaving the model-attributable agreement at less than 50%, or 35% in the baseline case. Age seems to be the most important such factor and region ranks a close second. Gender doesn’t matter at all, and ethnicity has mixed effects by ethnicity.

1. <http://www.name-prism.com> [↑](#footnote-ref-1)
2. See ./ 2-pager-udacity-classifiers-linkedin [↑](#footnote-ref-2)