Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
May 4, 2023
A Regional Conference of the Society for Industrial and Systems Engineering

Analyzing the Impact of Race on Special Forces Assessment and Selection

Mark Bobinski[1], James Starling[1], and Allison Brager[2]

[1]Department of Mathematical Sciences
United States Military Academy
West Point, New York 10996

[2]U.S. Army John F. Kennedy Special Warfare Center and School
Fort Bragg, North Carolina

Corresponding author's Email: mark.bobinski@westpoint.edu

Author Note: Sincerest gratitude to the U.S. Army John F. Kennedy Special Warfare Center and School for funding this research and providing necessary resources and data. Additional acknowledgement to MAJ Brager and LTC Starling for assistance and guidance throughout this research. The views expressed herein are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, or the Department of Defense.

Abstract: Special Forces Assessment and Selection (SFAS) is a series of physical, mental, and military challenges that test the readiness of a soldier for special operations missions. Over the last three fiscal years, white candidates have had a 26 percent pass rate compared to 20 percent for non-white candidates. This study uses logistic regression and decision tree models to determine if there is a significant difference in the passing rates for the two classifications of race, through a proportion test, and whether race is statistically significant to SFAS success. The proportion test showed that there was a significant difference between the passing rates of white and non-white candidates. Additionally, the outputs from each of the models proved that race is not statistically significant to passing SFAS.

Keywords: Military, Special Forces Assessment and Selection, Race, Predictive Modeling

1. Introduction

The John F. Kennedy Special Warfare Center and School (SWCS) is an organization within the United States Army that test, trains, and prepares Green Berets for the responsibilities of special forces operations. These soldiers are required to be fit, intelligent, and disciplined to handle the rigors of special forces. To become a member of special forces, soldiers must pass SFAS, a 24-day course that examines physical and mental ability and toughness (Army, 2022). This study aims to determine whether race is statistically significant to passing SFAS by classifying the candidates into a binary response variable for statistical modeling where the levels of the response variable are white and non-white races.

Throughout the research logistic regression and decision tree classification models will be used to determine the key characteristics of passing SFAS. The researchers will then examine the statistically significant variables and determine if there is a strong disparity among the two classifications of race within those variables. The aim of this research is to inform SWCS if there is any statistical evidence as to why the white population has a higher passing rate than the non-white population.

2. Literature Review

In a study conducted on Navy Sea, Air, and Land (SEAL) training, researchers looked to examine key characteristics of individuals who were successful within the course as well as create statistical models that identified relevant characteristics to predicting a soldier's success (Smith, Young, & Crum, 2020). The main model used throughout the study was a linear predictor model and one of the variables used within the research was the binary classification of race as white and non-white participants. The results of this study showed that the race variable was not statistically significant as it did not have a p-value less than 0.05. This applies to the current study with both the use of predictive modeling but also by classifying race into a binary variable with the two levels being white and non-white.

A non-military study looked to examine student's demographic attributes and determine if it had any contribution to their academic success in college (Batool et al., 2021). The researchers in this study used a random forest classification model to predict the success of students and determine the weight each variable had on the model. It is important to note that random forest models consist of multiple decision tree classifiers within the model. Additionally, this study used race and ethnicity as one of the explanatory variables. The researchers found that race had the fourth highest weight within the random forest model and that it accounted for 8.3 percent of the overall weight. The current study looks to use a decision tree classification model and a variable importance feature like the college study that will provide insight into the impact and weight that race has on SFAS results.

3. Methods

3.1 Data Cleaning

Since fiscal year 2020 the average passing rate for SFAS has been 25 percent. Of those who have attended the course, 87.7 percent classified their race as white and only 12.3 percent of people classified as other races. For the data set in this study, the white population had a total of 4,628 observations and the non-white population had a total of 646 observations. The white population has a passing rate of 26.2 percent while the other races have a passing rate of 20.4 percent.

Prior to statistical analysis, the data was gathered from SWCS in which case the 5,502 observations of 57 variables ranged from fiscal years 2020 to 2022. Each variable provided information on previous candidates to physical fitness test scores (*PT*), general technical test scores (*GT*), *age*, and time in the Army (*TIS*). New variables were then created based on the preexisting variables in the original data set. *season* was a binary variable representing the season of the year the candidate enrolled in the course. *rank_num* represented the numerical value associated with their rank. *Rank_below_sgt* was a binary variable for whether or not their rank was less than or equal to sergeant. *Region* is categorized as the region of the United States in which the soldier was coming from. *Comb_arms_bin* was a binary variable for whether the soldier was previously in a combat arms unit. Likewise, *infantry_bin* was binary for whether the soldier's previous unit was an infantry unit. *Higher_ed_bin* was binary for if a soldier attended any schooling beyond high school or equivalent to. *Dep_bin* was binary for if a soldier had any dependents. *Cb_ni* was binary for a soldier being in a combat arms unit but not being in an infantry unit. *AB_non_RGR* was binary for soldiers who had completed airborne school but not completed ranger school. *PT_above_avg* was binary for whether or not a soldier had scored at or above the average on their last physical fitness test. *PT_max* was binary for a soldier maxing their most recent physical fitness test. *Light_rs* is binary for a soldier coming from a unit that is considered light, meaning they operate primarily on the ground and with minimal technological support during operations. Finally, *big3* is a binary variable for if a soldier's recruiting station was either Fort Bragg, Fort Campbell, or Fort Benning.

The data set then dropped all observations with missing values and each of the numerical variables was normalized to a value between 0 and 1 which allowed the researchers to analyze the coefficients of the logistic regression model where each coefficient is equally comparable to one another.

The final step of the data cleaning process included cleaning the variable regarding the Army Physical Fitness Test (APFT). Since 1980 this test has been the consistent measurement criteria for soldiers in regards to their physical fitness (Holl, 2022). The scale of this test ranges from 0 to 300. However, in 2022, the Army implemented the Army Combat Fitness Test (ACFT) to replace the APFT. This new test is scored on a basis of 0 to 600 (Holl, 2022). The data set that was given to the researchers recorded all physical assessments under the category of APFT but some of the variables extended well beyond 300. Thus, it was assumed that the ACFT scores were recorded as APFT scores. These variables were then removed from the data set, leaving the final number of observations to be 5,271 over 37 different variables.

3.2 Methods of Analysis

The first step the researchers took was to use a proportion test to determine if the passing rates of the two classifications for race were significantly different. Then the researchers aimed to determine if race was statistically relevant to passing SFAS using a logistic regression model. In this model, all variables were included within the initial model, and the result of each candidate was the response variable. Automatic feature selection was used to include only variables that were statistically relevant to successfully passing the course. The researchers then recorded the variables included within the final model for further analysis.

Another logistic regression model was created using the same variables from the feature selection model mentioned above. However, this model included race as one of the explanatory variables. The reason being that this allows the researchers to compare a model with race and a model without race to see how well they perform in predicting SFAS success. In order to compare the models, the researchers used Akaike information criterion (AIC) which accounts for the goodness of fit for a model and the response variable (Portet, 2020). AIC acts as a scoring method to compare different models and determine which performs best in predicting the response variable. It is important to note that the lower the AIC score, the better the model performs. Using AIC allowed the researchers to compare the logistic regression models to determine whether the model including race performs better than the model excluding race, leaving all other explanatory variables consistent between the two.

A decision tree classification model was also configured in which case each of the variables was included within the model. Doing so allowed the researchers to then evaluate the feature importance within the model and determine which variables had the most significant weight and impact on predicting SFAS outcomes. The weights for feature importance have a total sum of one. Thus, all variables with a weight greater than 0.015 were considered, by the researchers, to be statistically significant and compared to the variables within the regression model.

Finally, the researchers then created two unique data sets, one consisting of only white candidates and one for non-white candidates. The mean for each of the variables was then calculated and used by the researchers to examine any significant differences among the two classifications of race, specifically within the variables appearing in both the logistic regression and decision tree models. If any variables proved to have large weights within these models, and also had a large difference in their percentage among the two races, then this could be a possible explanation for the overall difference in passing rates.

4. Results

The researchers first conducted a proportion test between the result and race variables to determine if there was a significant difference between the passing rate of white candidates and non-white candidates. The p-value of this test was 2.361e-05, implying that the difference in passing rates for the two races is statistically significant from one another. The two models used for this analysis were both logistic regression models. The first was configured through automatic feature selection in which case race was not in the final model, seen in Table 1a, showing that race is not statistically significant to passing selection. The model in which it was compared to was the same logistic regression model, except race was included as one of the explanatory variables, shown in Table 1b. To compare the two, the AIC was calculated for each model. AIC optimizes the features included within the model and penalizes models that have too many variables. Theoretically, if race was significant to passing SFAS, the AIC for a model would not increase because no penalization would occur. The AIC for the model without race was reported as 4245 and the model with race reported a 4247. Thus, because the model without race scored lower, it has a smaller prediction error on the response variable and is not tagged with the penalization of the AIC, meaning race had a negative impact on the model's ability to predict SFAS outcomes. One final note regarding the logistic regression models is that all variables contained a p-value less than 0.125. However, when race was added to the second model, it contained a p-value of 0.85 and a coefficient of -0.01 showing that it has very minimal impact on the likelihood of success.

Table 1. Summary Statistics for Logistic Regression without Race (a) and Logistic Regression with Race (b).

	Coefficient	P-value
(Intercept)	-13.5072	4.23E-16
class_month	0.39947	6.33E-05
RGR	1.06727	<2e-16
PLDC	-0.20266	0.111622
PT	3.87606	7.18E-11
TIS	-1.3915	1.55E-06
GT	1.59338	0.000357
SC	-6.4941	8.47E-05
FA	0.10441	5.35E-08
WVR1	-0.22688	0.001983
GENDER1	-0.36591	0.063741
rank_below_sgt	-0.49254	0.00022
comb_arms_bin	0.61039	6.61E-12
infantry_bin	-0.25989	0.002841
higher_ed_bin	-0.44934	3.23E-09
dep_bin	-0.12418	0.121792
AB_non_RGR	0.53979	8.10E-12
PT_max	0.28841	0.025712
light_rs	0.21095	0.005555

(a)

	Coefficient	P-value
(Intercept)	-13.53466	4.76E-16
class_month	0.3998	6.26E-05
RGR	1.06829	<2e-16
PLDC	-0.20325	0.110677
PT	3.87376	7.42E-11
TIS	-1.39319	1.52E-06
GT	1.59185	0.000362
SC	-6.52045	8.40E-05
FA	0.10489	6.06E-08
WVR	-0.22655	0.002019
GENDER	-0.36636	0.063415
rank_below_sgt	-0.4932	0.000217
comb_arms_bin	0.61174	6.95E-12
infantry_bin	-0.25898	0.002985
higher_ed_bin	-0.44858	3.61E-09
dep_bin	-0.12401	0.12233
AB_non_RGR	0.54063	8.12E-12
PT_max	0.28955	0.025309
light_rs	0.21041	0.005713
RACE	-0.01707	0.850013

(b)

The decision tree classification model was created and the weights for all variables were calculated. Race was placed as the 23rd highest weight of the 37 variables as shown in Figure 1. For reference, the weights of PT and GT, the two most relevant variables within the model, scored a 0.13 and 0.10 compared to that of 0.01 for race. Thus, based on the results of the study, the researchers have found that there is a statistically significant difference between the passing rates of white and non-white sol- diers using the proportion test. However, because race was not considered a statistically significant predictor in both the logistic regression and decision tree models, it then not considered a relevant factor in predicting success in SFAS. More importantly, race is not statistically significant enough to be considered a relevant factor in a candidate successfully navigating SFAS.

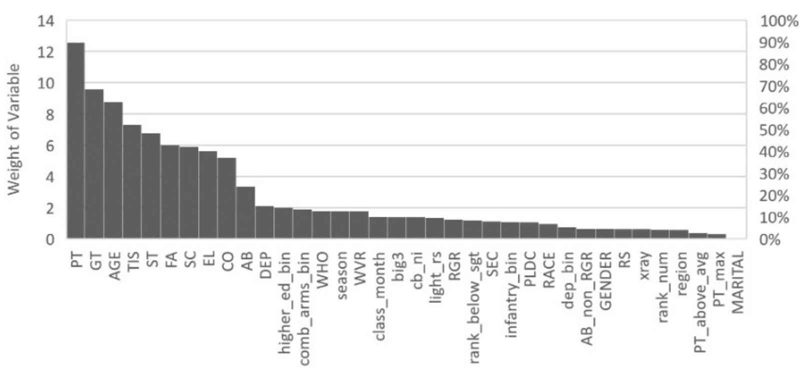

Figure 1. Feature Importance Scores for Decision Tree Model.

Figure 2 below shows the differences, in percentage, for the baseline statistics of the two racial categories. This was cal- culated by taking the average of each variable for the non-white population and subtracting that from the average of the white population. Variables in the graph marked with (*) correspond to a negative difference, meaning the non-white population had a higher percentage for that variable. It is important to note that the only variables considered for Figure 2 were those that had a percentage difference greater than 5. Of the variables determined to be statistically relevant in the models created

ISBN: 97819384964-4-8

throughout the study, none had a substantial difference for the two classifications of race.

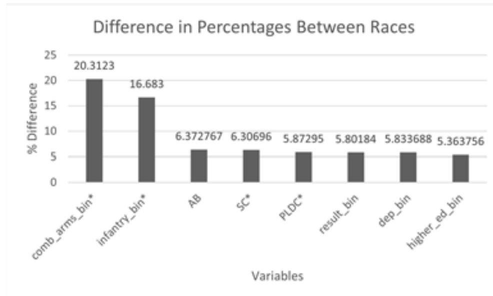

Figure 2. Difference in Baseline Statistics.

5. Implications and Future Studies

In analyzing the results of this study, the researchers were able to show that race has no significant impact on the result of SFAS. However, further analysis could be conducted with a larger sample size, in which a larger and more even distribution of observations for each race could be analyzed. Additionally, more variables could be recorded on each candidate prior to their arrival at SFAS. Statistics such as ruck march times, pull-ups, and other cognitive tests could prove to be statistically relevant and potentially provide further explanation to the difference in passing rates between the two races. Other analysis could be conducted on where majority of the different races were failing the course. Whether it be on physical examinations, cognitive tests, or military skills, there is potential to examine this difference further. One area that should be examined among the different phases that candidates could fail is the peer review process. Candidates who make it through the required training and tests are then required to rate their peers that endured the course with them. If candidates are rated, on average, significantly lower than their counter parts than they would be peered out and removed from the SFAS program. A significant difference between the two results of the peer review would be a compelling area of analysis for future research.

6. References

Army, U. (2022). *Special forces.* Retrieved 2023-01-31, from https://www.goarmy.com/careers-and-jobs/specialty-careers/special-ops/special-forces.html

Batool, S., Rashid, J., Nisar, M. W., Kim, J., Mahmood, T., & Hussain, A. (2021). A random forest students' performance prediction (rfspp) model based on students' demographic features. In *2021 mohammad ali jinnah university international conference on computing (majicc)* (pp. 1–4).

Holl, D. (2022). *Why fitness matters – reviewing history of army fitness testing.* Retrieved 2023-01-31, from https://www.army.mil/article/253358/why_fitness_matters_reviewing_history_of_army_fitness_testing

Portet, S. (2020). A primer on model selection using the akaike information criterion. *Infectious Disease Modelling*, *5*, 111–128.

Smith, E. N., Young, M. D., & Crum, A. J. (2020). Stress, mindsets, and success in navy seals special warfare training. *Frontiers in psychology*, *10*, 2962.