UNITED STATES MILITARY ACADEMY

WEST POINT, NEW YORK

HONORS THESIS

Predicting Cadet Candidates' Physical Success

by

John M. Stockdale

May, 2018

Thesis Advisors:	COL Joseph M. Lindquist
	CPT Dusty S. Turner
Second Reader:	MAJ Timothy J. Sikora

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** May/2018	**3. REPORT TYPE AND DATES COVERED** Senior Thesis (Honors)	
4. TITLE AND SUBTITLE Predicting Cadet Candidates' Physical Success		**5. FUNDING NUMBERS**	
6. AUTHOR(S) John M. Stockdale			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) United States Military Academy West Point, NY 10996-1786		**8. PERFORMING ORGANIZATION REPORT NUMBER**	
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A		**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**	
11. SUPPLEMENTARY NOTES			
12a. DISTRIBUTION / AVAILABILITY STATEMENT		**12b. DISTRIBUTION CODE**	

13. ABSTRACT (maximum 200 words)

The United States Military Academy is one of the most physically demanding universities in the world. Given its end-goal of commissioning Army Officers capable of meeting the physical demands of the military profession, the admissions office must appoint the most physically capable cadet candidates. This report attempts to help the admissions office tackle this problem and find which qualities of cadet candidates are most indicative of physical success at the Academy and subsequently in the Army. The approach to find physical success factors will use applied statistics in the form of linear and logistic regression. This will model which qualities of candidates historically proved to best predict physical success at the Academy. Given that the physical program of USMA is multi-faceted, the results varied per physical aspect of the Academy: extra-curricular involvement, work ethic, and raw general fitness were the best predictors of the composite physical program score (CPPS); high school sport involvement and success in those sports best predicted physical fitness test score (APFT) improvement over the four-years at the Academy; and both raw general fitness and success in high school sports best predicted success in the Academy's staple Indoor Obstacle Course Test (IOCT). Operating within the complexity of the admissions process, it is clear that a combination of high-school involvement, raw physical fitness, and success in high school athletics can help predict candidates best qualified for the US Military Academy's renowned physical program and a career in the US Army.

14. SUBJECT TERMS			**15. NUMBER OF PAGES** 20
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

PREDICTING CADET CANDIDATES' PHYSICAL SUCCESS

John M. Stockdale
CDT, FA
B.S., United States Military Academy, 2018

Submitted in partial fulfillment of the
requirements for the degree of

**BACHELOR OF SCIENCE IN OPERATIONS RESEARCH
WITH HONORS**

from the

**UNITED STATES MILITARY ACADEMY
May, 2018**

Author: John M. Stockdale

Approved by:

COL Joseph M. Lindquist
Thesis Advisor

CPT Dusty S. Turner
Co-Advisor

MAJ Timothy J. Sikora
Second Reader

Colonel Steven B Horton
Head, Department of Mathematical Sciences

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The United States Military Academy is one of the most physically demanding universities in the world. Given its end-goal of commissioning Army Officers capable of meeting the physical demands of the military profession, the admissions office must appoint the most physically capable cadet candidates. This report attempts to help the admissions office tackle this problem and find which qualities of cadet candidates are most indicative of physical success at the Academy and subsequently in the Army. The approach to find physical success factors will use applied statistics in the form of linear and logistic regression. This will model which qualities of candidates historically proved to best predict physical success at the Academy. Given that the physical program of USMA is multi-faceted, the results varied per physical aspect of the Academy: extra-curricular involvement, work ethic, and raw general fitness were the best predictors of the composite physical program score (CPPS); high school sport involvement and success in those sports best predicted physical fitness test score (APFT) improvement over the four-years at the Academy; and both raw general fitness and success in high school sports best predicted success in the Academy's staple Indoor Obstacle Course Test (IOCT). Operating within the complexity of the admissions process, it is clear that a combination of high-school involvement, raw physical fitness, and success in high school athletics can help predict candidates best qualified for the US Military Academy's renowned physical program and a career in the US Army.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

A special thanks to the United States Military Academy's Department of Physical Education and Admissions Office. First, the Department of Physical Education, COL Nicholas Gist in particular, helped advise this thesis and helped direct our project towards what he considers success in the USMA physical program. Second, the Admissions Offices shared the data required to conduct the statistical analysis and examination which allowed us to arrive at conclusions to our initial questions.

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

The United States Military Academy is always finding ways to improve. The Department of Physical Education mimics this tendency of the Academy's as it too is trying to encourage its cadets to constantly challenge and improve their physical fitness. However, alongside encouraging cadets to improve, the Department of Physical Education must also ensure that it is assessing and selecting the right cadets who are capable of physical success at the Academy. This introduces the role of the Military Academy's admissions office in selecting the right cadets. This report will analyze the existing practices of the admissions office and offer analysis and possible suggestions for improvement on the admissions process for selecting physically capable cadets. This report finds the best predictors for cadet physical success by conducting linear and logistic regressions for the cadet's physical score (CPPS), Indoor Obstacle Course Test (IOCT) score, and improvement on the Army Physical Fitness Test (APFT). The results show that the same few factors stayed consistent in predicting physical success, these are: Candidate Fitness Assessment score, the high school faculty's appraisal of the candidate, status as a recruited athlete, and extra-curricular involvement. However, although these predictors among many other measurements help select candidates who will be successful at the Academy, none acknowledge the role of functional fitness. The admissions office should consider implementing new tests or metrics to evaluate candidates' functional fitness. This would help the Department of Physical Education and the Academy fulfill its commitment to improvement.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

The United States Military Academy has the formidable challenge of accepting the most physically proficient candidates within the many legal constraints on the admissions process. The admissions office has to evaluate thousands of applicants and predict their physical success at the Academy and readiness for Officership using limited information such as high school athletic feats and one physical test. Notwithstanding, the Academy must choose the best candidates as long as they meet the requirements the law puts on the admissions process. These many challenges constrain the already daunting task of predicting athletic performance of thousands of candidates.

II. BACKGROUND

A. FITNESS DOMAINS

The United State Military Academy is not the only organization that has to predict athletic performance. Rather, many other institutions across the globe all have to try to tackle this challenge as well. Whether being another branch of the military under the Department of Defense, another physically demanding service organization, or simply an athletic team, many groups have tried to implement their own ways to forecast the physical success of individuals in their respective professions. Understandably, since the human body is widely understood in sciences such as sports medicine, many of the physical tests these organizations administer turn out to be relatively similar to each other. However, what makes these tests different from each other is the dissimilar end-goals and varying definitions of physical success of these organizations. The copious physical concerns that these organizations have can be grouped into three different aspects: general fitness, functional fitness, and body composition. All of these aspects are important and the admissions office should heavily considered them when evaluating physical ability. The most physically competent people, however, would excel in all three of these aspects.

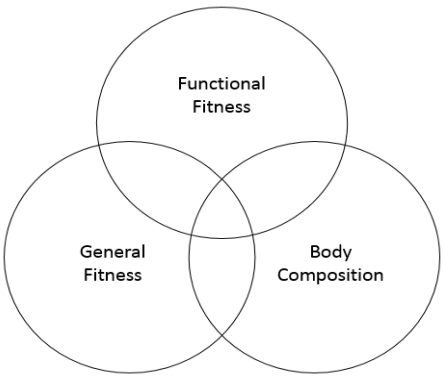

Figure 1. The Three Aspects of Fitness

B. ARMY'S DEFINITION OF PHYSICAL FITNESS

As a result, the Army organizes its training to consider all three fitness aspects including functional fitness, general fitness, and body composition. FM 7-22, the manual for Army Physical Readiness Training, defines physical readiness as a mixture of all of these aspects. In the same manual, the Army defines physical readiness as the combination of both physical fitness and health. Health in the Army's definition is the Soldier not being injured/sick and fully functioning in all physical regards. Physical fitness, on the other hand, is a combination of the Soldier being strong and having good endurance. In the context of Army Physical Readiness Training, its concern for health takes body composition and some aspects of functional fitness into consideration; and physical fitness takes functional and general fitness into consideration. Therefore, when choosing candidates to be future Army Officers, USMA must choose individuals who are proficient in all three of these aspects.

On a smaller level, the Army also has much more focused and exacting fitness measures as parts of physical readiness training. In accordance with FM 21-20, the manual for Physical Fitness Training, each Soldier must take a fitness test once every six months. This general fitness test measures how many push-ups a Soldier can do in two minutes, how many sit-ups they can do in two minutes, and how fast they can run two-miles. The Army has minimum standards for each event by age and gender and a soldier must pass all three events above the minimum standards in order to pass the test. This test, however, only measures muscular endurance and aerobic endurance and misses many other disciplines of fitness. The reason the Army does it this way is because this test must also be conducted while on deployments or training rotations where a lot of fitness equipment

will not be available. This test is rather inadequate in holistically evaluating Soldiers' fitness, but it suffices when concerning the many constraints put on Soldiers' schedules.

The Army is moving toward focusing on functional fitness, however. Because of the aforementioned inadequacy of the Army Physical Fitness Test, the Army wants to re-make the test with the goal of making it align more with the Army's mission. The aim of the proposed test is to examine the Soldier's ability to carry substantial weight, drag heavy objects, do short sprints, and throw. In other words, it wants to test the functional movements a soldier would have to perform in a combat or training scenario (Kimmons). Although the Army is still developing this test, it has made it clear enough to move to a more functional-focus when evaluating fitness.

Figure 2. Demonstration of the Army's Functional Fitness Exercises (Department of Physical Education).

The Army also has standards set for body composition. AR 600-9, the regulation for the Army's weight control program, sets minimum and maximum height and weight standards. A Soldier must weigh within a specified range for his/her age group, height, and gender. If

the Soldier doesn’t fall within that weight range, they then go under further evaluation of how much of that excess weight is fat and if they still can be considered healthy. If a Soldier is deficient in this, they are enrolled in the Army Body Composition Program (ABCP) and must lose a certain amount of weight over a given time, visit a nutritionist, and meet other demands in order to stay in the Army. If a Soldier is enrolled in ABCP and doesn’t meet its demands, that Soldier is separated from the Army under AR 600-9 regulations.

C. USMA'S ADMISSIONS PROCESS

When bringing cadets into the Academy, the United States Military Academy has many constraints to operate under. The primary law that limits physical considerations during the admissions process is Title 10 of the U.S. Code. Though this law is extensive in nature and touches on all aspects of the Military Academy, there are a few regulations in this code that in particular that make the most significant impacts on the admissions process. First, Title 10 Chapter 403 Section 4346 states that there must be 5 cadets at the Academy under nomination per each Congressman or Senator. This means that the Academy must give spots to certain candidates if they are filling a vacant congressional nomination slot in their district, regardless of how competitive they are relative to other candidates across the nation. Second, the Academy must admit applicants between that ages 17 and 24 while simultaneously meeting other demographic considerations. This regulation narrows the population size of the applicant pool substantially. In all, the Academy, Army, and nation have composition goals for the incoming classes of cadets which, as a result, keep the admissions office from solely accepting the most physically qualified cadets.

In addition to the regulations of the admissions process, there are a number of minimum requirements each candidate must meet in order to be considered qualified for admission to the Academy. First each candidate must pass a medical screening board (DoDMERB) in order to be considered medically qualified. Next, candidates must take a fitness test (Candidate Fitness Assessment—CFA) comprised of a minute of push-ups, a minute of sit-ups, pull-ups, a basketball throw, and a one-mile run. The candidates must meet a minimum standard score by age and sex in each event in order to be considered

physically qualified. Among all of these, the candidate must also be qualified by academic and leadership standards.

D. USMA'S DEFINITION OF FITNESS

Past all of the constraints of the Admissions board and the numerous qualifiers for cadet candidates, USMA has its own specific goals for fitness. The Department of Phjysical Education's mission statement is:

> "The Department of Physical Education develops warrior leaders of character who are physically fit and mentally tough by engaging cadets in activities that promote and enhance physical excellence."

To further unearth the perspective of the Department of Physical Education, I held a stakeholder meeting with the Head of the Department of Physical Education, Colonel Nicholas Gist. In it, COL Gist gave his perspective to what he looks for in cadets' physical fitness. He believes that all aspects of the Physical Program as a whole provide the best evaluation to a cadet's physical ability. A cadet, in his eyes, must be fit enough to pass the Army Physical Fitness Test, aggressive and adroit enough to pass boxing and submission grappling, comfortable enough in water to pass Survival Swimming, and athletic enough to pass Military Movement and intramural sports. However, COL Gist also said that of all of these, he thinks the Indoor Obstacle Course Test is the best mark of a cadet's fitness because it tests their strength, mobility, balance, coordination, willpower, and flexibility all at once and most closely replicates many tasks required of soldiers. He also made the observation that the predictors of most fit cadets are their pull-up score on the CFA and if they wrestled in high school. With all this said, USMA has a much more in-depth fitness program than the Army and it is imperative that the Admissions board selects the most physically fit cadets.

Figure 3. Cadet Candidates Being Evaluated on Pull-Ups (Association of Graduates).

III. GENERAL FITNESS

The first of the mentioned fitness aspects, general fitness, is the most widely used consideration of physical organizations when evaluating physical performance. General fitness, is the raw measurement of physical ability such as raw strength, speed, or endurance. It is used so commonly because it provides an explicit quantitative measurement of physical ability. Tests that look at general fitness are not subject to misinterpretation or error because they directly measure an individual's true current physical ability.

Most branches in the Department of Defense focus their physical tests towards measuring general fitness. The United States Air Force and its many schools test general fitness because it can tell the organization whether or not an individual will have the physical ability to complete a course and do the physical demands of the job (McGee). Other branches such as the Marine Corps, Navy, Coast Guard, and Army all have their own tests that measure general fitness. They all have some variation of running, strength, and endurance tests that are all aimed to gain a quantitative measure of an individual's physical ability. The Army's test, for example, is comprised of two-minutes of push-ups, two-minutes of sit-ups, and a timed two-mile run. These organizations can then use scores from their test to form a minimum standard for success for incoming recruits.

Figure 4. Cadet Candidates Being Evaluated on General Fitness (United States Military Academy Preparatory School).

Many sports organizations use general fitness tests because they provide a simple measurement of physical ability. However, most professional sports organizations take a much more in-depth and scientific approach to these tests. Rather than just testing speed, strength, or endurance, they will measure the body's mechanisms that drive these performance measures. For example, the Journal of Martial Arts reported that they tested individual athlete's VO2 Max, WATTMax, and HRMax to measure the wrestler's general fitness (Karjnincic). These measurements showed how well the wrestler's body will use oxygen when running, how well the wrestler's muscles will produce force, and the wrestler's overall heart rate. These ways of testing still measure raw fitness; however, they are more exacting because they are not subject to factors that can positively or negatively influence individual's scores (i.e. taller person running faster because of their longer stride).

IV. FUNCTIONAL FITNESS

The next aspect, functional fitness, takes an approach similar to general fitness. This category, when tested, can also quantitatively measure how fit an individual is. However, functional fitness does not just consider raw physical ability. It measures how well an individual can do these physical tests in a complex setting. Measuring functional fitness has all of the same benefits as measuring general fitness, except it is in the context of a multipart setting with the multifaceted physical challenges that mimic the physical tasks of the individual professions.

Many scientists have tested this component because it measures how well the body can synthesize all disciplines of fitness into doing one complex task. The International Journal of Sports Therapy reported that it tested individuals' ability to conduct the overhead deep squat because it incorporates muscular strength, balance, mobility, and endurance. Another test by the journal Trends in Sports Sciences measured individual's muscle elasticity and power by testing the distance they can cover in three consecutive hops. Tests like these evidently show that there is strength in understanding how well the human body can do several intricate tasks at once.

Figure 5. CrossFit Demonstrating Functional Fitness (Directly Fitness).

Service organizations, however, have used functional fitness tests in a much more direct measurement of athletic performance. The Dutch Police Department, for example, administered a test where officers had to do all of the physical tasks of the job as fast as possible (Strating). This test by the Dutch was a very clear illustration of how physically prepared an individual was for their job in the Police Department because it tested the exact physical demands of the job. In a similar way, the National Football League relies on past performance in football games and on players' abilities to do the tasks of the position in which they play (Lyons). As it was mentioned before, functional fitness tests can be great predictors of physical success because they test exactly what tasks the individual will be doing in that profession.

V. BODY COMPOSITION

The third category, body composition, covers the make-up of the individual's body and the factors that drive their physical performance. This component can be broken down to things such as an individual's flexibility, mobility, body-fat percentage, and other underlying mechanics of their body. This category is more of a qualitative measurement of their physical condition; however, it allows organizations to understand how durable the individual's body is for training and how much room it has for improvement.

Figure 6. Measuring Body Composition (Pak Health Shop).

Body composition has an integral part in physical performance. Body composition is what drives a person's ability to do work, to sustain work, and be overall effective in the physically demanding parts of the job. It is important to know what an individual's physical limits are, how much they can improve, and how resistant to injury they are. For example, the International Archives of Occupational and Environmental Health reported that they tested firefighters' body composition to predict how prone they will be to injury. The study showed that the firefighters' flexibility, grip strength, body fat percentage, and resting heart rate were the best predictors of injury resistance and longer careers in the business. Other organizations, especially competitive sports teams at elite levels, look for people with the optimal physique for their sport. The fitness aspect body composition reveals the

underlying make up of people and how that can transcend into physical ability into the future.

VI. DATA AND METHODOLOY

A. DATA

The admissions department stores records of all admitted candidates' application files and their corresponding grades and cadet records. With this, I was able to attain thousands of previous cadets' records and applications. However, operating within the constraints of the Institutional Review Board (IRB), I am only able to look at data from the USMA Class of 2008 or prior. Furthermore, I also am unable to incorporate the use of data that is too old because the data is either no longer on file, or the type of data stored has changed too much and is incomparable. With that, I have data from the 1998-2008 graduates, with a total of 13068 cadet records and applications on file.

The data is organized by each individual cadet. Each of the 13,081 cadet records has both pre-admissions candidate data and cadet data obtained from the individual's performance at the academy. The data for these cadets is comprised of administrative data, academic data, personal data, physical data, and military records. Furthermore, all of the data obtained for each cadet on file sums up to 725 specific pieces of information. In short, there are 13,081 rows of cadet files with 725 columns of individual data instances.

B. METHODOLOGY

In order to best define physical success as a cadet, after research and from COL Gist's input, I concluded that the best three ways to define success were the cadet's physical grade average (CPPS), their first-class IOCT performance, and their APFT improvement from plebe year to firstie year. I then use these three items as the data from which I will conduct analysis.

First, due to missing data, I would have to impute data. The best way, I found to impute the missing data, was to use the K-Nearest Neighbor method. From that, I would be able to conduct analysis and proper modeling on a complete data set. Next, I would then conduct linear regressions on each of the three aforementioned measures of cadet success—CPPS, firstie IOCT, and APFT improvement. Using step-wise multiple linear regression, I would find the best multiple linear regression model and its mathematical equation to predict each of the three response variables. Last, I would do the same but for logistic regression in order to have an equation to predict the probability of success for each of the three variables where each response variable is categorized as a success or failure.

After analysis, I will have 6 equations to best measure physical success as a cadet. I will have a linear regression equation to predict each a cadet's IOCT score, their APFT improvement, and CPPS using admissions numbers as the input variables. I will also have an equation to predict the probability of success for each of the three response variables again using admissions numbers as predictors.

VII. RESULTS

A. CPPS

The first approach to determining which variables are best for predicting physical success is through multiple linear regression. This method will provide an equation that will mathematically estimate what Cadet Physical Performance Score (CPPS) a candidate is most likely to earn based on his/her admissions numbers.

Using the statistical software, 'R,' I created the possible linear regression models. Using a package called 'leaps,' I was able to use statistical methods to find the best combination of predictor variables, being the one with the lowest AIC value, for regression equations containing 1 to 30 predictor variables. From there, I used residual plots, adjusted r-squared values, p-values, and more to determine which single equation had the most meaningful predictability.

Following the general multiple linear regression equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \cdots + \beta_n x_n + \varepsilon$$

The best statistical model was:

$$\hat{y} = -1.94 + 0.000188x_1 - 0.000189x_2 + 0.0809\sqrt{x_3} + 0.0704\sqrt{x_4} + 0.0122x_5 + 0.0709x_6$$

Where

$$\hat{y} = Predicted\ Response\ Variable$$

$$x_1 = Whole\ Candidate\ Score$$

$$x_2 = Extra\ Curricular\ Activity\ Score$$

$$x_3 = Factulty\ Appraisal\ Score$$

$$x_4 = PAE\ Score$$

$$x_5 = High\ School\ Sport\ Lettered\ Years$$

$$x_6 = ODIA\ Offers$$

With the following regression output:

```
Residuals:
     Min       1Q   Median       3Q      Max
-2.69188 -0.20039  0.03076  0.24475  1.22059

Coefficients:
                          Estimate  Std. Error t value             Pr(>|t|)
(Intercept)            -1.84335412  0.20358425  -9.055 < 0.0000000000000002 ***
whole_cand_scr.x        0.00018775  0.00001244  15.091 < 0.0000000000000002 ***
ex_cur_actvy_scr.y     -0.00018858  0.00004305  -4.381  0.00001198688144046 ***
sqrt(faculty_app_scr.y) 0.08092555  0.00787161  10.281 < 0.0000000000000002 ***
sqrt(pae_scr)           0.07035819  0.00290356  24.232 < 0.0000000000000002 ***
hs_sport_letter_yrs     0.01220428  0.00154388   7.905  0.0000000000000305 ***
odia_offer_num_types    0.07085371  0.01020637   6.942  0.0000000000417209 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3755 on 7823 degrees of freedom
Multiple R-squared:  0.1727,    Adjusted R-squared:  0.172
F-statistic: 272.1 on 6 and 7823 DF,  p-value: < 0.00000000000000022
```

Figure 7. CPPS Linear Regression Output

Notably, the multiple r-squared of this model is a small 0.172 which means the model only captures 17.2% of the variance within the data. Although this is a very low number, it does not invalidate the model. Rather than capturing the consistency within the data, this model highlights the most significant predictor variables for CPPS. Nonetheless, this model is still the most predictive out of all other possible multiple linear regression models.

To help explain this model, I included the following graphic. The bar chart below shows the proportionate contributions to CPPS by each individual predictor variable.

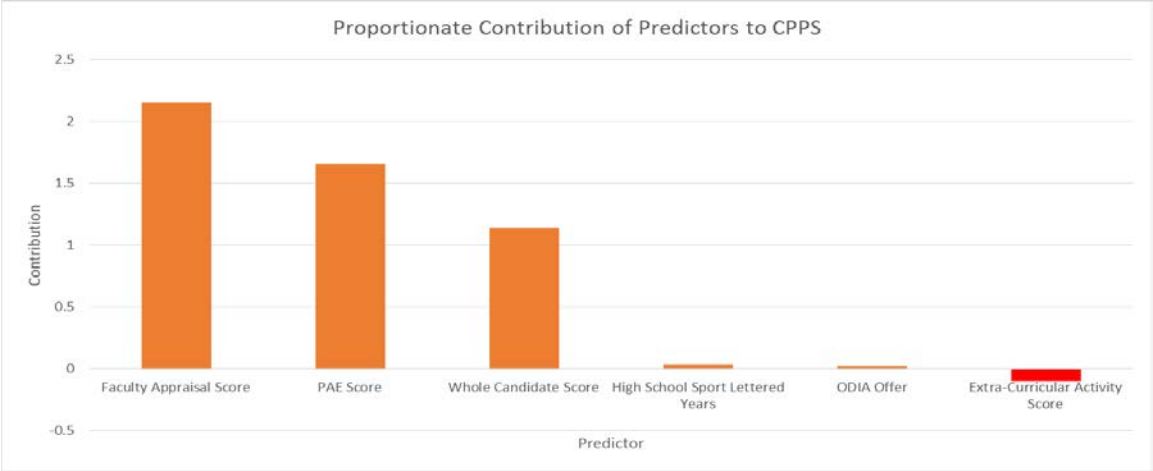

Figure 8. CPPS Proportionate Contribution Bar Chart

It is clear that the student's faculty appraisal, physical fitness test score, and whole composite application are the best indicators to his/her future physical program score at USMA.

There still is a fair amount of error within this model. There is some, but not much, non-linearity in the residual for each individual predictor. However, there is nothing notable enough to invalidate its ability to have real-world application.

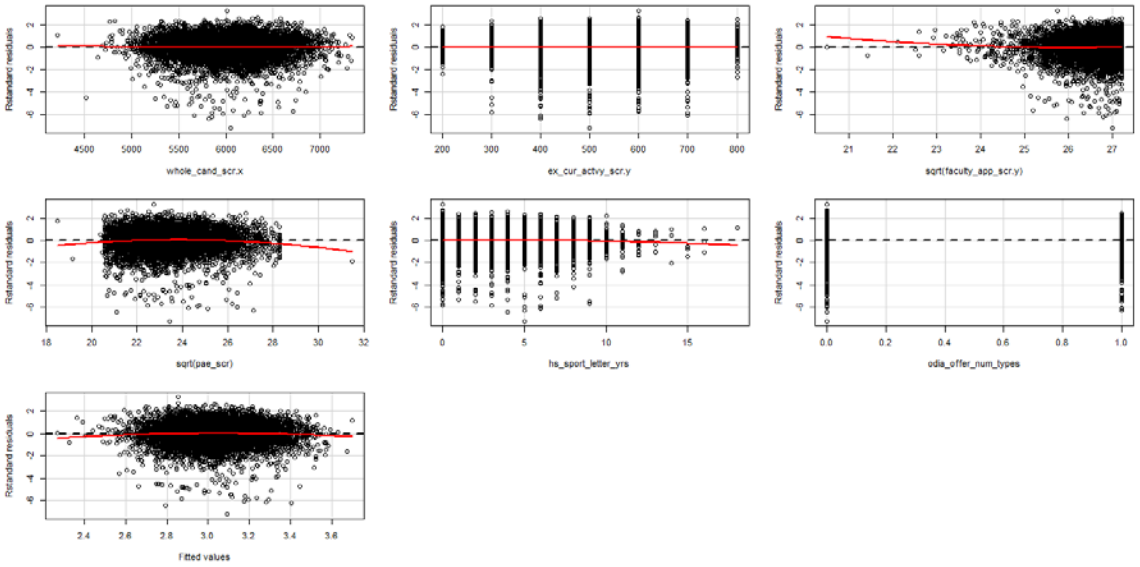

Figure 9. CPPS Residual Plots

The second method for determining how to best predict CPPS will be through logistic regression. Similar to linear regression, logistic regression will produce an equation that can model how successful a cadet will be with regard to their physical program score. Logistic regression will predict the probability of meeting a 'successful' score. In this case, a successful score would be above a 3.2 CPPS which is about one standard deviation above the mean for each USMA class which is what COL Gist and myself considered to be successful.

Similar to the linear regression, I will use both analytic methods, statistical inference, and common sense applications to the real world to find the best model with the least error.

If the general multiple logistic regression function is:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$$

The model had different predictors than the linear regression model for CPPS.

$$\ln\left(\frac{p}{1-p}\right) = -14.5 + 0.0678x_1 + 0.00784x_2 - 0.00134x_3 + 0.00704x_4 + 0.000845x_5$$

Where

$$\frac{p}{1-p} = Predicted\ Probability\ of\ Success$$

$$x_1 = High\ School\ Sport\ Lettered\ Years$$

$$x_2 = PAE\ Score$$

$$x_3 = Extra\ Curricular\ Activity\ Score$$

$$x_4 = Faculty\ Appraisal\ Score$$

$$x_5 = Whole\ Candidate\ Score$$

With regression output:

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1363  -0.4081  -0.3216  -0.2507   2.9050

Coefficients:
                       Estimate  Std. Error z value             Pr(>|z|)
(Intercept)          -12.7228735   1.1913762 -10.679 < 0.0000000000000002 ***
hs_sport_letter_yrs    0.0665355   0.0133516   4.983      0.0000006249655 ***
odia_offer_num_types   0.4088347   0.0913285   4.477      0.0000075866716 ***
pae_scr                0.0071458   0.0005435  13.148 < 0.0000000000000002 ***
ceer_scr.x             0.0049833   0.0007347   6.783      0.0000000000118 ***
faculty_app_scr.x      0.0037454   0.0016848   2.223                0.0262 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4891.2  on 9784  degrees of freedom
Residual deviance: 4585.2  on 9779  degrees of freedom
AIC: 4597.2

Number of Fisher Scoring iterations: 6
```

Figure 10. CPPS Logistic Regression Output

Through both models, it is clear that there are 3 strong predictors. Physical Performance Score is best predicted by raw fitness test scores, faculty appraisal, and the strength of the candidate's entire application packet.

B. FIRSTIE IOCT

Similar to CPPS, the first approach to predicting the candidate's Firstie IOCT score was through multiple linear regression.

Following the general multiple linear regression equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \cdots + \beta_n x_n + \varepsilon$$

The best equation was:

$$\hat{y} = -1647 + 12.7\sqrt{x_1} + 16.7\sqrt{x_2} + 511.3Log\sqrt{x_3} + 5.03x_5 + 11.06x_5$$

Where

$$\hat{y} = Predicted\ Response\ Variable$$

$$x_1 = CEER\ Score$$

$$x_2 = Faculty\ Appraisal\ Score$$

$$x_3 = PAE\ Score$$

$$x_4 = High\ School\ Sport\ Captain\ Years$$

$$x_5 = ODIA\ Offers$$

With the following regression output:

```
Residuals:
    Min      1Q  Median      3Q     Max
-474.35  -38.00   11.92   65.74  324.49

Coefficients:
                       Estimate Std. Error t value             Pr(>|t|)
(Intercept)           -1647.404     91.635 -17.978 < 0.0000000000000002 ***
sqrt(ceer_scr.x)         12.702      1.168  10.878 < 0.0000000000000002 ***
sqrt(faculty_app_scr.x)  16.700      2.439   6.847      0.00000000000809 ***
log(sqrt(pae_scr))      511.297     21.857  23.393 < 0.0000000000000002 ***
hs_sport_cpt_yrs          5.025      1.204   4.174      0.00003023751546 ***
odia_offer_num_types     11.064      3.278   3.375              0.000741 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 119.7 on 7824 degrees of freedom
Multiple R-squared:  0.09912,   Adjusted R-squared:  0.09855
F-statistic: 172.2 on 5 and 7824 DF,  p-value: < 0.00000000000000022
```

Figure 11. Firstie IOCT Score Linear Regression Output

The resulting model also has a very low r-squared of 0.099 and is less than assuring in predictive value. But this too is just for highlighting the best combination of predictors and this model is the best of all potential models.

Again, the bar chart below shows the proportionate contributions to Firstie IOCT score by each individual predictor variable.

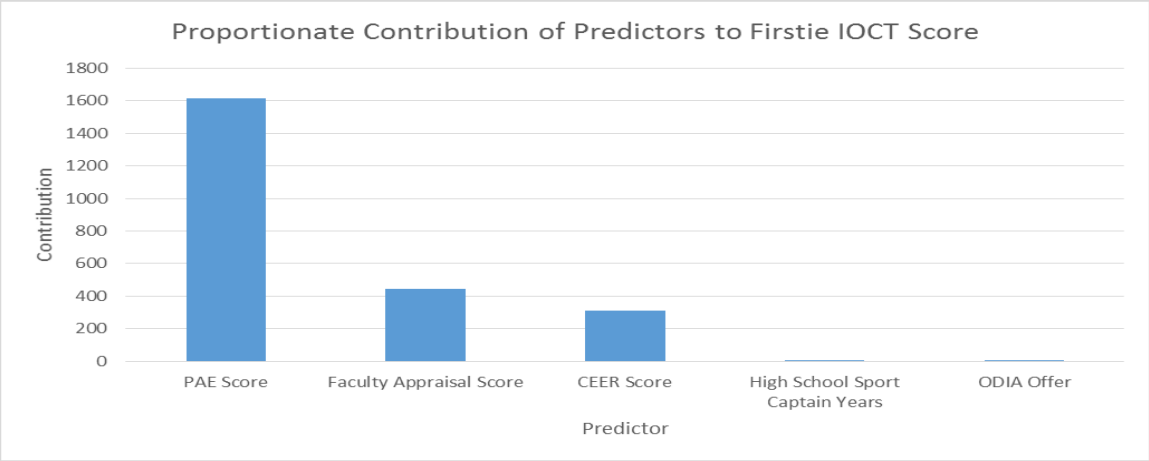

Figure 12. Firstie IOCT Score Proportionate Contribution Bar Chart

It is clear that the student's PAE Score, faculty appraisal, and CEER score are the most powerful predictors of their IOCT score. Notably, PAE score has the most significant contribution to their score during their Firstie year USMA.

Similar to CPPS, there still is a fair amount of error within this model. There is some, but not much, non-linearity in the residual for each individual predictor. However, there is nothing notable enough to invalidate its real-world application.

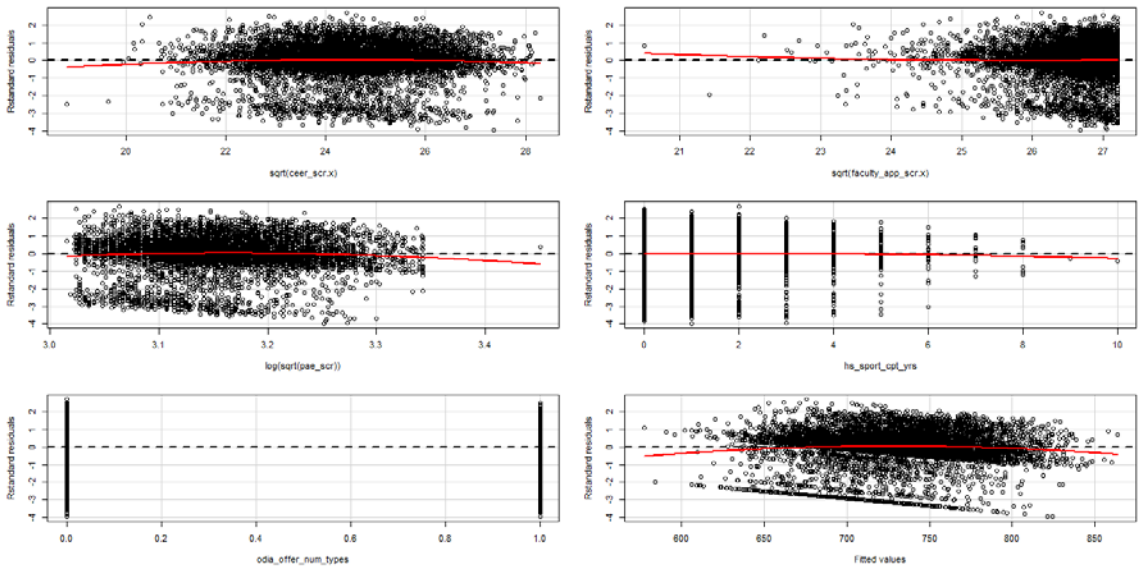

Figure 13. Firstie IOCT Score Residual Plots

The second method for determining how to best predict Firstie IOCT scores will be through logistic regression as well. I will make the bench mark of success a score of 900, which is a 'tabbing' score at USMA. This score will differentiate cadets who mastered the obstacle course from those who still have improving to do.

If, again, the general multiple logistic regression function is:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$$

The final logistic model had different predictors than the linear regression model for the IOCT scores.

$$\ln\left(\frac{p}{1-p}\right) = -12.7 + 0.0665x_1 + 0.4084x_2 + 0.00714x_3 + 0.00498x_4 + 0.00375x_5$$

Where

$$\frac{p}{1-p} = Predicted\ Probability\ of\ Success$$

$$x_1 = High\ School\ Sport\ Lettered\ Years$$

$$x_2 = ODIA\ Offers$$

$$x_3 = PAE\ Score$$

$$x_4 = CEER\ Score$$

$$x_5 = Faculty\ Appraisal\ Score$$

With regression output:

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1363  -0.4081  -0.3216  -0.2507   2.9050

Coefficients:
                       Estimate  Std. Error z value             Pr(>|z|)
(Intercept)          -12.7228735   1.1913762 -10.679 < 0.0000000000000002 ***
hs_sport_letter_yrs    0.0665355   0.0133516   4.983      0.0000006249655 ***
odia_offer_num_types   0.4088347   0.0913285   4.477      0.0000075866716 ***
pae_scr                0.0071458   0.0005435  13.148 < 0.0000000000000002 ***
ceer_scr.x             0.0049833   0.0007347   6.783      0.0000000000118 ***
faculty_app_scr.x      0.0037454   0.0016848   2.223               0.0262 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4891.2  on 9784  degrees of freedom
Residual deviance: 4585.2  on 9779  degrees of freedom
AIC: 4597.2

Number of Fisher Scoring iterations: 6
```

Figure 14. Firstie IOCT Score Logistic Regression Output

Through both models, we can infer that the candidate's raw fitness test score and success in high school sports are the best two predictors for success in the IOCT during their Firstie year at USMA.

C. APFT IMPROVEMENT

Just as for Firstie IOCT score CPPS, the first approach to predicting the candidate's improvement on the Army Physical Fitness Test (APFT) between the spring of Firstie year and fall of plebe year was through multiple linear regression.

As always, following the general multiple linear regression equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \cdots + \beta_n x_n + \varepsilon$$

The best equation was:

$$\hat{y} = -39.5 + 0.0902 x_1 - 0.610 x_2 + 7.27 x_3$$

Where

$$\hat{y} = Predicted\ Response\ Variable$$

$$x_1 = CEER\ Score$$

$$x_2 = High\ School\ Sports\ Years$$

$$x_3 = ODIA\ Offers$$

With the following regression output:

```
Residuals:
     Min       1Q   Median       3Q      Max
-132.232  -20.268   -1.886   20.129  165.107

Coefficients:
                    Estimate Std. Error t value             Pr(>|t|)
(Intercept)       -39.489376   5.803666  -6.804      0.0000000000118 ***
ceer_scr.x          0.090226   0.009239   9.766 < 0.0000000000000002 ***
hs_sport_TOTAL_yrs -0.610174   0.150633  -4.051      0.0000521019826 ***
odia_offer_num_types  7.279915   1.329868   5.474      0.0000000468699 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 32.62 on 3711 degrees of freedom
Multiple R-squared:  0.03156,   Adjusted R-squared:  0.03078
F-statistic: 40.31 on 3 and 3711 DF,  p-value: < 0.00000000000000022
```

Figure 15. APFT Improvement Linear Regression Output

Of the three linear regression models, this model has the lowest r-squared at 0.032 and explains very little variation in APFT improvement. Despite the low predictive ability of the model, it highlights the best combination of predictors for APFT improvement.

The bar chart below shows the proportionate contributions to APFT improvement by each individual predictor variable.

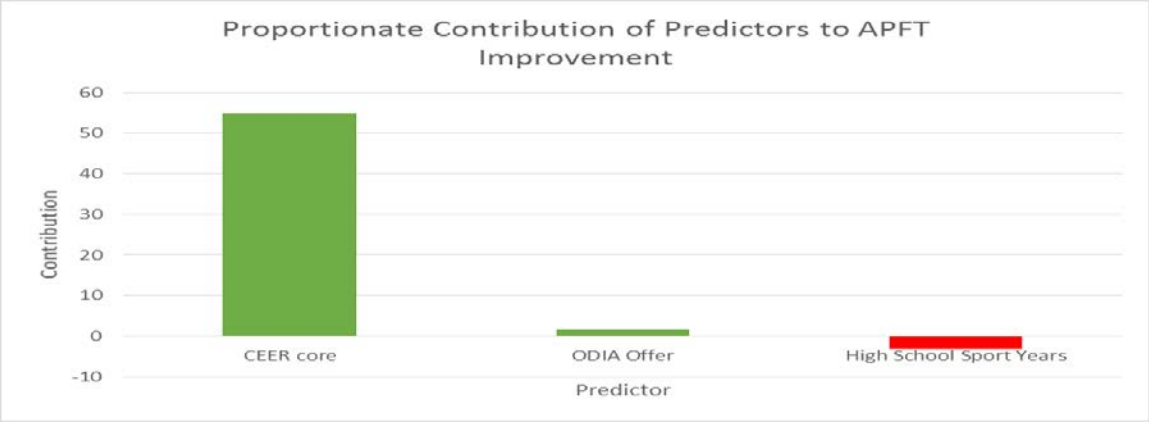

Figure 16. APFT Improvement Proportionate Contribution Bar Chart

It is clear that the student's CEER score us by far the most impactful predictor of their future APFT improvement. Notably, CEER score is a composite score for the candidate's school record, so this does not hold much realistic value aside from possibly being correlated to his/her work ethic. Again, ODIA offers has some contribution which makes sense provided that these collegiate athletes would be guaranteed to be exercising every day.

Similar to the other two response variables, there still is a fair amount of error within this APFT improvement model. Of note, no transformations were necessary. There is some, but not much, non-linearity in the residual for each individual predictor. However, there is nothing notable enough to invalidate any real-world application.

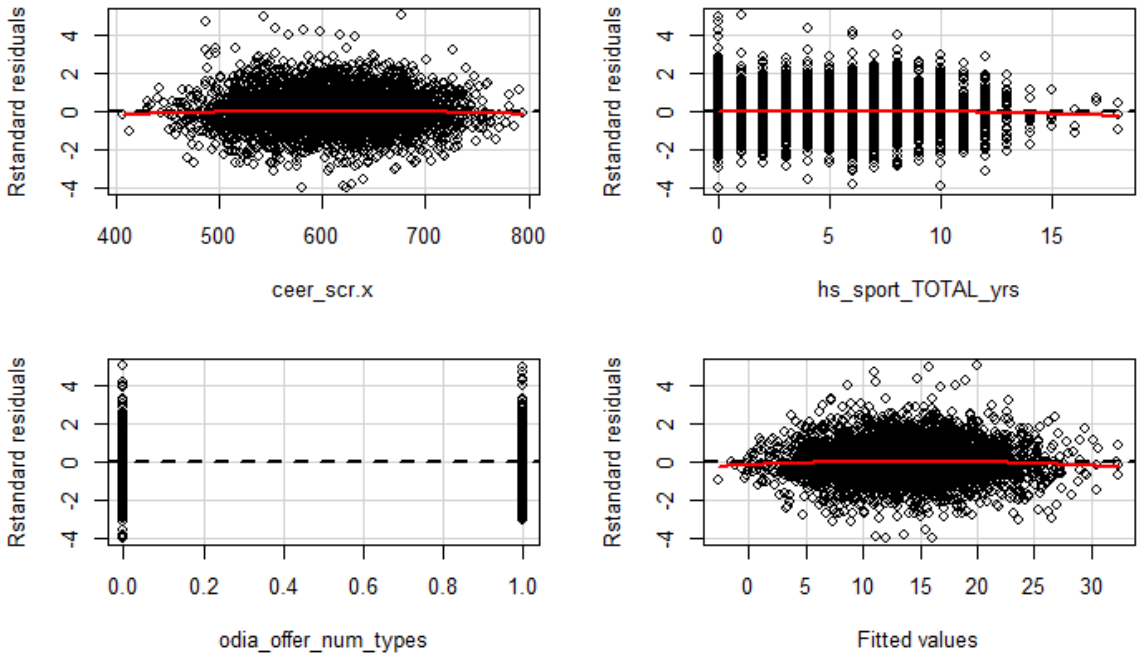

Figure 17. APFT Improvement Residual Plots

The second method for determining how to best predict APFT improvement will be through logistic regression as well. I will make the bench mark of success either an improvement of over 10 points on the APFT and/or a starting score of 300 their plebe

year. This will allow cadets who start with the highest possible score to be considered successful, given that they are unable to score any higher.

If the general multiple logistic regression function is:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$$

The logistic model had different predictors than the linear regression model for the APFT improvement.

$$\ln\left(\frac{p}{1-p}\right) = -3.6 + 0.00587x_1 + 0.000684x_2 - 0.000441x_3 - 0.0260x_4 + 0.298x_5$$

Where

$$\frac{p}{1-p} = Predicted\ Probability\ of\ Success$$

$$x_1 = CEER\ Score$$

$$x_2 = Athletic\ Activity\ Score$$

$$x_3 = Extra\ Curriculuar\ Activity\ Score$$

$$x_4 = High\ School\ Sport\ Years$$

$$x_5 = ODIA\ Offers$$

With regression output:

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7506  -1.1802   0.8807   1.1290   1.5730

Coefficients:
                    Estimate Std. Error z value             Pr(>|z|)
(Intercept)        -3.6031241  0.4737147  -7.606   0.0000000000000282 ***
ceer_scr.x          0.0058662  0.0005996   9.783 < 0.0000000000000002 ***
athl_actvy_scr.x    0.0006847  0.0003623   1.890              0.058780 .
ex_cur_actvy_scr.y -0.0004413  0.0003292  -1.340              0.180135
hs_sport_TOTAL_yrs -0.0259670  0.0099147  -2.619              0.008818 **
odia_offer_num_types  0.2984721  0.0904861   3.299              0.000972 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5143.3  on 3714  degrees of freedom
Residual deviance: 5033.7  on 3709  degrees of freedom
AIC: 5045.7

Number of Fisher Scoring iterations: 4
```

Figure 18. APFT Improvement Logistic Regression Output
Both models seem very different from each other and raise a few concerns.

However, we can conclude through these models that the common trends for APFT improvement are sport involvement at the collegiate level and the candidate's proven work-ethic.

VIII. CONCLUSION

The admissions process at the United States Military Academy is a very delicate and challenging process. Especially daunting is the task of choosing candidates who will be physically fit for the Academy and the Army profession. However, when attempting to meet the Army's physical standards and the wants of the Department of Physical Education, there were valuable measures that indicate future physical success in the IOCT, CPPS, and improvement on the APFT.

The results were understandable and realistic in nature. Raw physical aptitude and a candidate's entire application best predicted success for the entire USMA physical program. Raw physical aptitude best predicted success in the Indoor Obstacle Course Test. The candidate's holistic academic success and success in organized sports best predict improvement in the Army Physical Fitness Test.

As a caveat, these results were not entirely predictive in nature. Each regression model had an R-squared value of less than 0.2 which is alarmingly low. Although these models still best summarized the trends with data, they did not carry over to being predictive for future observations.

When tying this all back to research done on measuring physical aptitude, these models encompass the general fitness and body composition components. However, within the data, there was no useful data that covered the functional fitness component. Moving forward, it would be in the best interest of the Army and the Military Academy to focus on the aforementioned best predictors of physical success, but also open up metrics for the functional fitness component for the admissions office to consider. However, although these predictors help select candidates who will be successful at the

Academy, none acknowledge the role of functional fitness. The admissions office should consider implementing new tests or metrics to evaluate candidates' functional fitness.

IX. WORKS CITED

Association of Graduates. “USMA Class of 2018 Reception Day.” Video. 9 Jul 2014. https://www.youtube.com/watch?v=dknQ_qxxr28. 20 Oct 2017.

Cesar, Guilherme M., et al. "Prediction of Athletic Performance of Male and Female Athletes Measured by Triple Hop for Distance." Trends in Sport Sciences, vol. 24, no. 1, Mar. 2017, pp. 19-25. EBSCOhost, 0-search.ebscohost.com.usmalibrary.usma.edu/login.aspx?direct=true&db=s3h&AN=122301111&site=eds-live&scope=site.

Clifton, Daniel R., et al. "Overhead Deep Squat Performance Predicts Functional Movement Screen™ Score." International Journal of Sports Physical Therapy, vol. 10, no. 5, Oct. 2015, pp. 622-627. EBSCOhost, 0-search.ebscohost.com.usmalibrary.usma.edu/login.aspx?direct=true&db=s3h&AN=112626291&site=eds-live&scope=site.

Department of Physical Education. Picture of me hurdling over a tire. United States Military Academy. Taken November 2015.

Directly Fitness. “CrossFit Training Pros and Cons: A Closer Looks” Image. 07 Jul 2017. http://www.directlyfitness.com/store/crossfit-training-pros-and-cons/

Headquarters Department of the Army. “Physical Fitness Training.” FM 21-20. Washington D.C. 1998.

Headquarters Department of the Army. “Army Physical Readiness Training.” FM 7-22. Washington D.C. Oct. 2012.

Headquarters Department of the Army. “The Army Weight Control Program.” AR 600-9. Washington D.C. Nov. 2006.

Karnincic, Hrvoje, et al. "Factors of Success in Advanced Level Wrestling; Reliability and Validity of Several Diagnostic Methods." Journal of Combat Sports & Martial Arts, vol. 6, no. 2, July 2015, pp. 77-83. EBSCOhost, 0-search.ebscohost.com.usmalibrary.usma.edu/login.aspx?direct=true&db=s3h&AN=122208871&site=eds-live&scope=site.

Kimmons, Sean. “With six events, the Army Combat Readiness Test aims to replace APFT, cut injuries.” Army News Service. 18 Sep. 2017. Web. https://www.army.mil/article/193953/with_six_events_new_army_combat_readiness_test_aims_to_replace_apft_cut_injuries

Lyons, Brian D., et al. "On the Predictive Efficiency of past Performance and Physical Ability: The Case of the National Football League." Human Performance, vol. 24, no. 2, Apr. 2011, pp. 158-172. EBSCOhost, 0-search.ebscohost.com.usmalibrary.usma.edu/login.aspx?direct=true&db=s3h&AN=60106989&site=eds-live&scope=site.

McGee, John Christopher, et al. "Physical Screening Predictors for Success in Completing Air Force Phase II Air Liaison Officer Aptitude Assessment." Military Medicine, vol. 182, 2017 Supplement, pp. 281-286. EBSCOhost, doi:10.7205/MILMED-D-16-00199.

Pak Health Shop. “Weight to Scale Large Display.” Image. 2016. www.pakhealthshop.com/shop/weight-scale-to-measure-large-display/

Poplin, Gerald, et al. "Fire Fit: Assessing Comprehensive Fitness and Injury Risk in the Fire Service." International Archives of Occupational & Environmental Health, vol. 89, no. 2, Feb. 2016, pp. 251-259. EBSCOhost, doi:10.1007/s00420-015-1068-4.

Strating, M., et al. "A Job-Related Fitness Test for the Dutch Police." Occupational Medicine, vol. 60, no. 4, June 2010, pp. 255-260. EBSCOhost, 0-search.ebscohost.com.usmalibrary.usma.edu/login.aspx?direct=true&db=eih&AN=51207921&site=eds-live&scope=site.

United States Military Academy Preparatory School. “Fitness Curriculum.” Image. https://www.usma.edu/usmaps/SitePages/Physical%20Fitness%20Curriculum.aspx. 20 Oct 2017.

U.S. Code. Title 10. Subtitle B. Part III. Chapter 403. “The United States Military Academy.” Washington D.C. 2016.

X. Appendix

p(success)	CEER Score	Athletic Activity Score	Extra Curricular Activity Score	HS Sport Total Years	ODIA Offer
0.83448	H	H	M	M	H
0.803574	H	H	M	M	M
0.789056	H	H	M	M	L
0.816748	H	M	M	M	H
0.783391	H	M	M	M	M
0.767812	H	M	M	M	L
0.769749	H	L	M	M	H
0.730659	H	L	M	M	M
0.712678	H	L	M	M	L
0.609325	M	H	M	M	H
0.558616	M	H	M	M	M
0.536437	M	H	M	M	L
0.579624	M	M	M	M	H
0.528046	M	M	M	M	M
0.505689	M	M	M	M	L
0.508413	M	L	M	M	H
0.456293	M	L	M	M	M
0.43418	M	L	M	M	L
0.325466	L	H	M	M	H
0.281366	L	H	M	M	M
0.26362	L	H	M	M	L
0.299012	L	M	M	M	H
0.25713	L	M	M	M	M
0.240401	L	M	M	M	L
0.242397	L	L	M	M	H
0.206114	L	L	M	M	M
0.191847	L	L	M	M	L
	Key				
	CEER Score	Athletic Activity Score	Extra Curricular Activity Score	HS Sport Total Years	ODIA Offer
	H=800	H=800	M=530	M=3.25	H=1
	M=600	H=620			M=0.3
	L=400	L=200			L=0

Figure 1: APFT Improvement Logistic Regression Probability Output Chart

p(success)	HS Sport Letter Years	ODIA Offer	PAE Score	CEER Score	Faculty Appraisal Score
0.798792	M	M	H	H	H
0.77392	M	M	H	H	M
0.544968	M	M	H	H	L
0.596644	M	M	H	M	H
0.560532	M	M	H	M	M
0.308552	M	M	H	M	L
0.351047	M	M	H	L	H
0.318077	M	M	H	L	M
0.140295	M	M	H	L	L
0.146935	M	M	M	H	H
0.129316	M	M	M	H	M
0.049395	M	M	M	H	L
0.060307	M	M	M	M	H
0.052437	M	M	M	M	M
0.018993	M	M	M	M	L
0.022932	M	M	M	L	H
0.019836	M	M	M	L	M
0.00703	M	M	M	L	L
0.036751	M	M	L	H	H
0.031851	M	M	L	H	M
0.011379	M	M	L	H	L
0.014017	M	M	L	M	H
0.01211	M	M	L	M	M
0.00427	M	M	L	M	L
0.005172	M	M	L	L	H
0.004463	M	M	L	L	M
0.001566	M	M	L	L	L
	Key				
	HS Sport Letter Years	ODIA Offer	PAE Score	CEER Score	Faculty Appraisal Score
	M=2	M=0.3	H=975	H=800	H=740
			M=550	M=600	M=700
			L=350	L=400	L=420

Figure 2: Firstie IOCT Logistic Regression Probability Output Chart

p(success)	HS Sport Letter Years	PAE Score	Extra Curricular Activity	Faculty Appraisal Score	Whole Candidate Score
0.98399	M	H	M	H	H
0.952891	M	H	M	H	M
0.894603	M	H	M	H	L
0.978951	M	H	M	M	H
0.938675	M	H	M	M	M
0.865284	M	H	M	M	L
0.865761	M	H	M	L	H
0.679752	M	H	M	L	M
0.471093	M	H	M	L	L
0.662468	M	M	M	H	H
0.392446	M	M	M	H	M
0.213252	M	M	M	H	L
0.597619	M	M	M	M	H
0.328319	M	M	M	M	M
0.170203	M	M	M	M	L
0.170783	M	M	M	L	H
0.06348	M	M	M	L	M
0.027657	M	M	M	L	L
0.272849	M	L	M	H	H
0.109919	M	L	M	H	M
0.049268	M	L	M	H	L
0.221151	M	L	M	M	H
0.085464	M	L	M	M	M
0.037735	M	L	M	M	L
0.037884	M	L	M	L	H
0.012793	M	L	M	L	M
0.005409	M	L	M	L	L
	Key				
	HS Sport Letter Years	PAE Score	Extra Curricular Activity	Faculty Appraisal Score	Whole Candidate Score
	M=2	H=975	M=530	H=740	H=7350
		M=550		M=700	M=6000
		L=350		L=420	L=5000

Figure 3: CPPS Logistic Regression Probability Output Chart