Blue Crow Sports Group – Manager, Sports Science Data Application

Alex Kikos

ASSESSMENT	METRICS	EXPLAINATION
30-15 IFT: Link 1	Final Speed (km/h)	The speed of the final stage completed before being removed from the test.
	V02max	An estimation of the VO2max value for the athlete based on the final speed of the test, age, body weight and gender.
0-30Y Linear Speed: Link 1	0-10Y Split	Time to complete the first 10Y in seconds
	10-30Y Split	Time to complete the first 20Y in seconds
	0-30Y Split	Time to complete the total distance from 0 to 30Y in seconds
Anthropometry	7 Site Fat %	Overall body % percentage taken using skinfold methods from seven sites
	Weight (lb)	Body weight in pounds
	Height (in)	Player's height in inches
CMJ (Force Plates): Link 1	Concentric Impulse	The force generated during triple extension (upward movement phase)
	Concentric Impulse – 100ms	The force generated during triple extension (upward movement phase) in 100ms
	Concentric Mean Power / Body Weight	The quotient of a subject's average power output during the concentric (or upward movement/triple extension) phase of the CMJ and the subject's weight
	Eccentric Mean Power / Body Weight	The quotient of a subject's average power output during the eccentric (or from the start of the subject's descend and lowest point during a squat before jumping up) phase of the CMJ and the subject's weight
	RSI - Modified	A version of RSI that can be measured by calculating the flight time to jump contraction time ratio
	Vertical Velocity @ Take off	The speed at which a vertical jump is done by the athlete
Max Strength (Trapbar Deadlift): Link 1 Link 2	Relative Max Strength	The value of the estimated 1RM divided by the body weight of the athlete
	Estimated 1RM	The estimated 1RM based on weight and number of repetitions using Brzycki's method

Tasks

1. The team is using force plates to evaluate the neuromuscular characteristics of the players using a CMJ. Please answer the following questions:

a. Provide a definition for each of the metrics collected. As much as possible use simple terms like if you were explaining it to a coach.

Please see above

b. What are your thoughts on the overall metrics collected within the context of this team (amateur football)? Would recommend removing or adding any metrics moving forward? Please explain the rationale behind your answer.

These are great tests and metrics to have on a team and serve as a strong foundation to understanding performance. Additional variables that could be considered are player load from devices such as Catapult, as well as acceleration data to better understand signal attenuation during a given session/practice/game, which can help with injury prevention. Ground reaction force (GRF) would also be a beneficial metric to collect on individuals, specifically footballers, as the amount of force they apply on every step or touch on the pitch can influence injury risk and can also provide insight to the power at which they are able to generate.

2. Some of the assessments in the data contain multiple trials for each test. For example, the linear running speed or the CMJ test have multiple trials. Please develop a solution to report the results from tests with multiple trials. How do you calculate the results, how do you decide when to remove or keep a trial? Work with the dataset and develop a solution to report only one result per assessment. The new data set must be used to complete questions 3 and 4.

Depending on who the S/C person I am working and what they are interested in seeing, there are two ways I would approach this. Option 1 would be to utilize an athlete's peak or max score from a given trial, as it indicates their present ceiling for a given assessment. For example, MLB scouting departments prefer to use max exit velocity (EV) when referring to a player's EV, rather than his average, as it shows what he is capable of. The second option, and my recommendation, would be to take the mean value from all trials to account for some outliers and provide a more realistic indication of where an athlete is truly at. This would be done by grouping the athlete, test and metrics of interest, along with the test date and taking the average value for that date (lines 38-40 in R Shinny app).

3. The director of strength & conditioning for Blue Crow Sports is interested to know whether 30Y sprint times can be predicted with the data that we have available and he/she is asking if you can help. How do you approach this question analytically? Please elaborate a solution and provide both the code and process that you follow to come up with an answer as well as an explanation of your results.

There are two ways I would approach the prediction of 30Y Sprint Times, depending on if the S/C director wanted a model to plug values in (linear regression) or have a prediction model & visual output (regression tree) that they can easily see what the biggest factors are that influence the outcome. A linear regression model would allow the S/C director (or anyone) to view the corresponding weights of a given assessment/metric and apply that coefficient to every test score an athlete has in order to predict their 30Y sprint time. A regression tree is also an optimal supervised data mining technique to predict 30Y sprint times while also quantifying the weight/importance of each assessment/metric the athlete has done. A sample model of this can be found in lines 65-121 in R Shiny app. I applied data management techniques outside of R that turned each unique metric value into a column value and used the value from the original dataset to be placed in the corresponding column/row (line 78 in R Shiny app). Additionally, since blank/NA values existed in every row for this method, those values were turned into 0s for the purposes of running a clean regression tree. I understand that this is not a recommend data management technique and only complete classes from the original dataset should be used in future projects. The pruned tree for the regression model can be found in Figure 1 in the Appendix.

4. Blue Crow would like to have an in-house application to report this information and ensure that coaches are able to access it on a daily basis. Please create an interactive application that contains the following functionalities:

a. A tab that enables users to select an athlete and see the most recent results for all tests, shown as a percentile. Provide some context such as "good", "average", "poor", "excellent", etc.

Please run R Shiny app

b. A second tab that enables users to filter by position, test and metric and display a table or a chart to show the reference values from all the historical data

Please run R Shiny app

c. A third tab that lets users filter by player, test, metric, date and display a longitudinal trendline to track changes over time. Try to provide some context, for example, how does the coach know when there is a meaningful improvement?

Please run R Shiny app

Appendix

Figure 1: Pruned Regression Tree for finding 0-30y values using all indicators