Embargoed until 9 a.m. (EDT) August 12, 2015

STATISTICAL MODEL PREDICTS WITH HIGH ACCURACY PLAY-CALLING TENDENCY OF NFL TEAMS

SEATTLE, WA, AUGUST 12, 2015 – If a defensive coordinator of a National Football League (NFL) team could predict with high accuracy whether their team's opponent will call a pass or run play during a game, he would become a rock star in the league and soon be a head coach candidate.

William Burton, an industrial engineering student who is minoring in statistics at North Carolina State University (NCSU), and collaborator Michael Dickey, a statistics major who graduated from NCSU in May, have built a statistical model that predicts the play-calling tendency of NFL teams with high accuracy. Burton today presented the model at a session during the 2015 Joint Statistical Meetings (JSM 2015) in Seattle.

Their model, which correctly called run and pass plays at a high rate when tested using play-by-play data from actual NFL games, could be used by casual fans and even NFL defensive coordinators during real games to predict their opponent's next play.

"A valuable skill for NFL coaches is to be able to anticipate whether the opposing team will call a pass or run play. If the offensive play type can be predicted—say a pass—the defensive coordinator can call a blitz or coverage play to gain an advantage," explained Burton during his presentation.

Burton and Dickey used 2000 through 2014 NFL play-by-play data from ArmChair Analysis to conduct an initial analysis of the probability of a pass in a NFL game. This analysis revealed that pass probability in NFL games has risen by more than 2 percentage points from 54.4% during the 2000 season to 56.7% during the 2014 season. Armed with this information, they determined the model should be developed using data from the 2011–2014 seasons.

Next, they had to decide which factors most influence an offensive team's play selection. These include yards to go, the play down (first, second, third or fourth), time remaining, point differential, offensive points, defensive points, interaction between yards to go and down, cumulative number of fumbles, cumulative number of interceptions, field position, timeouts remaining for the offense, timeouts remaining for the defense and yards gained on the previous plays. They considered many other variables, such as play lag (what had occurred on the previous play) and current weather conditions (precipitation/wind speed), but found they did not have a significant effect on play-calling.

Burton and Dickey then developed logistic regression and random forest models using the ArmChair Analysis play-by-play data seasons to predict future play types. While building the logistic regression model, they determined separate models needed to be created for each quarter of a game because the behavior of the selected variables change by quarter. For example, if a team is losing in the fourth quarter, it is much more likely to throw a pass, while the winning team is less likely to call a pass play. Conversely, in the first quarter, point differential has no benefit to predicting play type.

Each quarter has its own quirks that are not picked up if modeled together. As a result, six unique logistic regression models were created—one each for the first quarter, second quarter, third quarter, fourth quarter winning, fourth quarter losing and fourth quarter tied.

To test their model, Burton and Dickey randomly selected 20 games from completed NFL seasons. The model's best result was correctly predicting 91.6% of plays in a 2014 game between the Jacksonville Jaguars and Dallas Cowboys, with the average prediction accuracy over all 20 games 75%.

The following is a list of five games with the highest prediction accuracy rates from the 20 tested. (Note: Only pass or runs plays are included; punts and field goal attempts are not included.)

2014 Dallas Cowboys at Jacksonville Jaguars

- Total # of Plays: 119
- Total # of Plays Correctly Predicted: 109
- Total # of Plays Incorrectly Predicted: 10
- Percent of Plays Correctly Predicted: 91.6%

2013 Baltimore Ravens at Denver Broncos

- Total # of Plays: 148
- Total # of Plays Correctly Predicted: 134
- Total # of Plays Incorrectly Predicted: 14
- Percent of Plays Correctly Predicted: 90.5%

2011 New York Giants at New England Patriots

- Total # of Plays: 128
- Total # of Plays Correctly Predicted: 109
- Total # of Plays Incorrectly Predicted: 19
- Percent of Plays Correctly Predicted: 85.16%

2014 New England Patriots at Seattle Seahawks (Super Bowl XLIX)

- Total # of Plays: 121
- Total # of Plays Correctly Predicted: 96
- Total # of Plays Incorrectly Predicted: 25
- Percent of Plays Correctly Predicted: 79.33%

2013 Arizona Cardinals at St. Louis Rams

- Total # of Plays: 107
- Total # of Plays Correctly Predicted: 84
- Total # of Plays Incorrectly Predicted: 23
- Percent of Plays Correctly Predicted: 78.50%

Now that the final model has been constructed and successfully tested, Burton and Dickey have created an interactive visualization using the R Shiny package. This visualization is an intuitive and easily interpretable tool individuals—from the casual fan to perhaps NFL coaches—can use to make real-time decisions based on current game conditions, explained Burton during his presentation.

JSM 2015 is being held August 8–13 at the Washington Convention Center in Seattle. More than 6,000 statisticians—representing academia, business and industry, as well as national, state and local governments—from numerous countries are attending North America's largest statistical science gathering.

About JSM 2015

JSM, which has been held annually since 1974, is being conducted jointly this year by the American Statistical Association, International Biometric Society (ENAR and WNAR), Institute of Mathematical Statistics, Statistical Society of Canada, International Chinese Statistical Association, International Indian Statistical Association, Korean International Statistical Society, International Society for Bayesian Analysis, Royal Statistical Society, and International Statistical Institute. JSM activities include oral presentations, panel sessions, poster presentations, professional development courses, an exhibit hall, a career service, society and section business meetings, committee meetings, social activities and networking opportunities. Click here for more information about JSM 2015.

About the American Statistical Association
The ASA is the world's largest community of statisticians and the second-oldest continuously operating professional society in the United States. Its members serve in industry, government and academia in more than 90 countries, advancing research and promoting sound statistical practice to inform public policy and improve human welfare. For additional information, please visit the ASA website at www.amstat.org.

<div align="center">###</div>

For more information:
Jeffrey A. Myers
Office: (703) 684-1221, Ext. 1865
Mobile: (540) 623-7777
E-mail: Jeffrey@amstat.org