Correlation of team statistics with performance in the NFL

Daniel Loman* and Matthew Bellis†
Siena College
(Dated: December 5, 2013)

abstract

INTRODUCTION

The NFL is a constantly evolving league. Since 1970 passing and rushing statistics have succomb to change from both rule changes and incoming trends. The following graph shows pass and rush yard data from 1970 to 2012.

From the graph you cannot dispute that the league has become a passing league. Before the 1978 season rules were implemented to improve both the passing attack and player safety. Since then the league has been subject to more subtle rule changes and passing trends, like the spread offense. Due to these causes NFL passing stats have skyrocketed in recent years while leaving the running game in the dust. Because the game is ruled by the passing offense, there exists the stigma that a good team must possess an elite quarterback to succeed in the NFL. The phrase "defense wins championships" is uttered far less and the running game has been extremely devalued. I decided to take a look correlations between passing yards, rushing yards, defensive passing yards, defensive rushing yards, and win percentage.

CORRELATION COEFFICIENTS

The correlation coefficient is measure of linear correlation between two variables, and is given by equation (1).

$$r = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}\sqrt{\sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2}} \qquad (1)$$

It ranges between -1 and 1, with 1 representing a perfect correlation, 0 representing no correlation, and -1 representing a perfect inverse correlation.

ESTIMATES OF THE UNCERTAINTIES

Of course, the correlation coefficient itself tells us nothing without the uncertainty. The bootstrap method is a computational method that uses resampling to find the uncertainty of the correlation coefficient, and was introduced in 1979 by Dr. Bradley Efron with his paper Bootstrap method: Another Look at the Jackknife. The bootstrap method involves creating several new data pairs the same size as the original data pair by randomly selecting elements from the original data pair. Each new data pair has a new correlation coefficient, which can be put into a separate array containing all the new correlation coefficients. This array will distribute close to a normal gaussian curve, and should peak around the correlation of the original pair. A predetermined range of the new correlation coefficients represents the uncertainty for the original. In this exercise I used an uncertainty ranges of 1 standard deviation and 95

RESULTS

The correlation coefficients.

INTERPRETATION OF RESULTS

Can we draw any conclusions?

CONCLUSIONS

Summary of the analysis.