Analysis of Thursday Night NFL Winning Margins

Timothy S. Vaughan
University of Wisconsin – Eau Claire

Key Words: *t*-test, Time series, Statistical process control.

Abstract

This paper introduces a dataset and associated analysis of the scores of National Football League (NFL) games over the 2012, 2013, and first five weeks of the 2014 season. In the face of current media attention to "lopsided" scores in Thursday night games in the early part of the 2014 season, *t*-test results indicate no statistically significant difference between the winning margins in Sunday games vs. Thursday games during the 2012 and 2013 seasons. Interestingly, there is a statistically significant difference between the Sunday vs. Thursday game margins over the first five weeks of the 2014 season. Moreover, statistical process control methods suggest an "out of control" condition for Thursday night game margins during the first five weeks of the 2014 season. The exercise provides students with an opportunity to apply a variety of hypothesis testing and graphical analysis tools to a question of current interest in the popular media.

1. Introduction

Instructors of statistical methods courses generally recognize the value of cases and exercises analyzing real data (e.g., Carter, Felton, and Schwertman 2014), particularly when the data address an issue of current concern in the popular media. Such datasets tend to motivate greater student interest in the analysis, and demonstrate practical relevance of the statistical methods at hand. Moreover, presenting a dataset with reference to an open question to be answered, as opposed to simply an opportunity to apply a particular statistical tool, demonstrates how various statistical tools can be used in unison, each one contributing additional insight into the question at hand.

In the case study presented below, students are able to apply a variety of informal, exploratory graphical analysis tools as well as basic hypotheses tests of statistical significance. The analysis

produced a somewhat surprising result, leaving open the very real question as to the extent to which examining the past helps to predict the future. While the results presented here were generated using the Data Analysis Toolpak in Microsoft Excel, clearly the analysis could be conducted using any statistical software of choice.

1.1 Background

The NFL has scheduled a varying number of Thursday night games since the 2006 season. The Thursday night games are generally viewed as motivated by a desire for additional television air time, beyond the traditional Sunday games and the highly-billed Monday night game, without interfering with high school and college games traditionally played on Friday and Saturday.

On October 2, 2014, the Green Bay Packers defeated the Minnesota Vikings by a score of 42 to 10, after leading 42-0 at the start of the fourth quarter. This Thursday night win added fuel to an existing controversy over "lopsided" scores in the Thursday night game sequence. Indeed, the first five weeks of the 2014 season produced the Thursday night scores shown in Table 1 below:

Table 1. Thursday Night NFL Scores, Weeks 1-5 of the 2014 season

Week	Visitor		Home Team		Winning Margin
1	Packers	16	Seahawks	36	20
2	Steelers	6	Ravens	26	20
3	Buccaneers	14	Falcons	56	42
4	Giants	45	Redskins	14	31
5	Vikings	10	Packers	42	32

NFL blogs, television and radio talk shows, and the popular press continued their critical analysis of the Thursday night games, proposing alternative explanations for why the games have such large margins, and offering alternative explanations for how to "fix" Thursday night football (see, for example, Ruiz 2014 and Schalter 2014). Some on-air personalities ultimately questioned whether the Thursday night game sequence should be discontinued.

The trained statistician, of course, has a natural tendency to question whether such casual observation represents a statistically significant event, as opposed to the common tendency toward over-reaction to random variation. Toward this end, the dataset described below was collected for all NFL regular season games played during the 2012 and 2013 seasons, as well as the first five weeks of the 2014 season. (This is the last week of games played as of the writing of this paper.) The dataset provides an intriguing exercise for the introductory statistics course, and can be used to motivate a variety of issues in statistical analysis.

> *Teaching Hint: Start a discussion of the Thursday night scores shown above, generating speculation as to why the winning margins might be larger on Thursday vs. Sunday, prior to raising the possibility that the situation could simply be an example of failing to recognize random variation.*

2. The Dataset

NFL scores are readily available from a number of websites. To construct the dataset, all regular season scores (excluding pre-season and playoff games) from the 2012, 2013, and first five weeks of the 2014 season were retrieved from www.nfl.com/scores. The data set and documentation file can be found at:
http://www.amstat.org/publications/jse/v23n1/vaughan/NFL_Scores_by_Day.csv and
http://www.amstat.org/publications/jse/v23n1/vaughan/NFL_Scores_by_Day_documentation.txt.

The dataset contains 588 records with the following fields:
Season   {2012, 2013, 2014}
Week   Week number within the season the game was played
Day   Day of week on which the game was played {Wed, Thu, Sat, Sun, Mon}
Visitor Team   Name of the visiting (away) team
Visitor Team Score
Home Team   Name of the home team
Home Team Score
Margin   Absolute difference between the two team scores

Football fans may have alternative definitions of "lopsided" and "blowout" games, some of which are not completely characterized by the final score of the game. Nonetheless, we proceed with the assumption that an underlying pattern of "lopsided" or "blowout" games will, across a large sample, be manifested by larger average winning margins as reflected in the final score. Summary statistics for the final winning margin across the 588 records are shown in Table 2 below:

Table 2. Summary Statistics for Winning Margins

Day	n	Minimum	Mean	Maximum	Variance	Std. Dev.
Sun	512	0	11.82	58	93.94	9.69
Mon	38	1	10.79	28	67.63	8.22
Wed	1	7	7.00	7	-----	-----
Thu	36	2	14.69	42	104.96	10.25
Sat	1	13	13.00	13	-----	-----
Overall	588	0	11.92	58	92.91	9.64

The histogram shown in Figure 1 demonstrates a skewed distribution for the game margins, as might be expected due to the lower bound at zero.

Figure 1. Histogram of Winning Margins, All Data
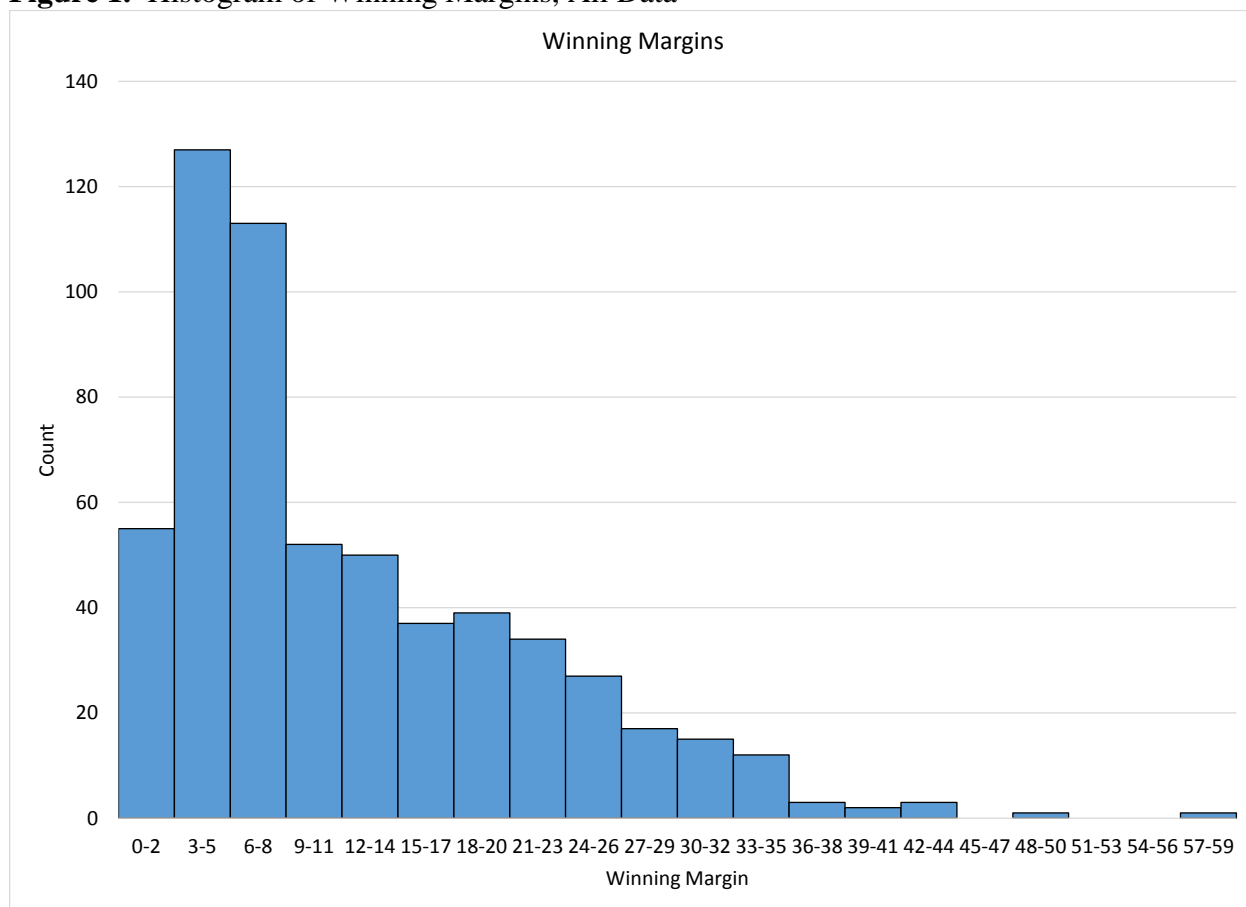
Teaching Hint: Ask the students to predict the shape of the histogram prior to showing it to them, or prior to having them develop it. Upon review, point out that a symmetric distribution has "enough room" on both sides of the mode for the frequency to taper off in both directions. Distributions that are bounded below at zero and unbounded above frequently end up highly skewed. This helps to develop student intuition about the data they are analyzing, as opposed to following a purely "mechanical" approach. This observation also leads to subsequent discussion of the t-test's robustness to non-normality, and the motivation for applying a non-parametric test, both of which are discussed below.

3. Analysis

While the dataset included a number of Monday night games, and a very small number of Wednesday and Saturday games, the current controversy primarily suggests a comparison of Thursday night vs. Sunday game margins. Toward that end, the first, most direct analysis prompted by the question at hand is a simple t-test comparing the winning margin for Thursday night vs. Sunday games. Our initial investigation involved a comparison of game margins across the entire dataset. Defining μ_S as the mean winning margin for Sunday NFL games, and μ_T as the mean winning margin for Thursday night NFL games, we tested:

H_0: $\mu_T = \mu_S$
H_1: $\mu_T > \mu_S$

The output from the Excel Data Analysis ToolPak is displayed in Table 3. (The standard deviation has been added to this and all subsequent reports to simplify interpretation.)

Table 3. *t*-test results for Thursday vs. Sunday night game margins, entire dataset.

t-Test: Two-Sample Assuming Equal Variances		
All Seasons, Winning Margin	*Thursday*	*Sunday*
Mean	14.69	11.82
Variance	104.96	93.94
Standard Deviation	10.25	9.69
Observations	36	512
Pooled Variance	94.65	
Hypothesized Mean Difference	0	
df	546	
t Stat	1.7157	
P(T<=t) one-tail	0.0434	
t Critical one-tail	1.6476	
P(T<=t) two-tail	0.0868	
t Critical two-tail	1.9643	

Given the one-sided alternative hypothesis, the one-tailed *p*-value of .0434 indicates a statistically significant difference at the α = .05 significance level. This result was a bit surprising, as the dataset was collected with the expectation of demonstrating typical "media overreaction" to random variation.

> *Teaching Hint: Discuss the use of the one-tailed vs. two-tailed hypothesis test. In this case the question of interest was prompted by claims that Thursday night game margins are "too large," and (by implication) larger than those of the traditional Sunday games. The question at hand would lead us to determine, <u>a priori</u>, that we will reject our null hypothesis only if the average Thursday margin is sufficiently <u>greater</u> than the average Sunday margin, leading to a one-sided alternative hypothesis. (We are not interested in demonstrating Sunday margins are significantly smaller than Thursday margins. An average Sunday margin less than the average Thursday margin, by any amount, would be grounds to fail to reject our null hypothesis.) As such, Type I error can only occur in the case of the Thursday sample mean being greater than the Sunday sample mean, and proper evaluation of the Type I error probability includes only the upper tail of the t distribution.*

3.1 Analysis of the 2014 Season

We subsequently questioned the statistical significance of the five Thursday night margins from the 2014 season, which are the real focus of the current controversy. Repeating the analysis for only the first five weeks of the 2014 season produced the results shown in Table 4 below:

Table 4. *t*-test results for Thursday vs. Sunday night game margins, first five weeks of the 2014 season (equal variances assumption)

t-Test: Two-Sample Assuming Equal Variances		
2014 Season, Winning Margin	*Thursday*	*Sunday*
Mean	29.00	12.08
Variance	86.00	69.45
Standard Deviation	9.27	8.33
Observations	5	65
Pooled Variance	70.42	
Hypothesized Mean Difference	0	
df	68	
t Stat	4.3453	
P(T<=t) one-tail	2.378E-05	
t Critical one-tail	1.6676	
P(T<=t) two-tail	4.755E-05	
t Critical two-tail	1.9955	

Teaching Hint: Tables 2 and 3 above report very similar sample variances for the Sunday and Thursday night games. Indeed, the F-test for unequal variances (not presented here) produces a p-value of 0.30, failing to demonstrate a statistically significant difference between the two sample variances. (Of course, failing to reject the null hypotheses H_0: $\sigma^2_T = \sigma^2_S$ does not prove the null hypothesis to be true, although the F test result in this context is frequently given that interpretation. Depending on the level of the course, this analysis could be presented.)

The variances presented in Table 4 above similarly fail to demonstrate a statistically significant difference (F-test p-value is again .30). However with the small sample size (n=5) for Thursday night games we have little power to detect unequal variances. For that reason it is prudent to additionally test the difference between the two sample means using the more conservative (less powerful) unequal variances t test.

Teaching Hint: Aside from the issue of detecting unequal variances as discussed above, the small sample size (n=5) only becomes problematic when we <u>fail</u> to demonstrate a statistically significant difference. This result could easily be due to low power associated with a small sample size. A larger sample, while certainly desirable in the

interest of power, is not inherently "required" for conducting valid t tests. We simply need to remember that failing to reject a null hypothesis never proves the null hypothesis to be true.

The more conservative unequal variances *t*-test produced the results shown in Table 5.

Table 5. *t*-test results for Thursday vs. Sunday night game margins, first five weeks of the 2014 season (unequal variances)

t-Test: Two-Sample Assuming Unequal Variances

2014 Season, Winning Margin	Thursday	Sunday
Mean	29.00	12.08
Variance	86.00	69.45
Standard Deviation	9.27	8.33
Observations	5	65
Hypothesized Mean Difference	0	
df	5	
t Stat	3.9594	
P(T<=t) one-tail	0.0054	
t Critical one-tail	2.0150	
P(T<=t) two-tail	0.0107	
t Critical two-tail	2.5706	

Interestingly, again referring to the one-tailed *p*-values, the *t* tests suggest there <u>is</u> a statistically significant difference between the Sunday vs. Thursday night NFL winning margins during the first five weeks of the 2014 season.

Teaching Hint: While the significant result actually was unexpected, in retrospect it was very useful to frame the analysis up to this point as an exercise in demonstrating over-reaction to random variation. Instead, this turned out to be an excellent demonstration of the need to <u>test our preconceptions against data</u>. (If our preconceptions were always correct, there would be no need to bother with analysis of data!)

Teaching Hint: Discuss the non-normality of the data as shown in Figure 1, in the context of the t-test being reasonably "robust" to non-normality. In too many cases, the test assumptions (which arise from the assumptions underlying mathematical derivation of the test statistic distribution) come across to students as a set of rigid requirements that must be met, otherwise the test "does not work."

In contrast, students should understand, and be able to critically evaluate inevitable departures from the assumptions underlying the test statistic derivation. ("Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful" – G.E.P. Box). In the present case, both distributions (Thursday and Sunday margins) would be expected to be skewed in the same direction as shown in Figure 1. This condition is not considered to have an invalidating impact on the Type I error probability now <u>approximated</u> using the t distribution. Nonetheless, a nonparametric

test, in the face of the small sample size and non-normality of the data, is discussed below.

3.2 Analysis of the 2012 and 2013 Seasons

This result naturally prompts the question as to whether this difference has existed in previous seasons. Examining the 2012 and 2013 seasons in combination (as well as individually) demonstrated no statistically significant difference. The result for the combined 2012 and 2013 seasons is shown in Table 6.

Table 6. *t*-test results for Thursday vs. Sunday night game margins, 2012-2013 seasons

t-Test: Two-Sample Assuming Equal Variances		
2012-2013, Winning Margin	*Thursday*	*Sunday*
Mean	12.39	11.78
Variance	71.38	97.66
Standard Deviation	8.45	9.88
Observations	31.00	447.00
Pooled Variance	96.00	
Hypothesized Mean Difference	0	
df	476	
t Stat	0.3344	
P(T<=t) one-tail	0.3691	
t Critical one-tail	1.6481	
P(T<=t) two-tail	0.7382	
t Critical two-tail	1.9650	

3.3 Graphical Analysis

It is typically useful, particularly in introductory statistics classes, to have students examine graphical evidence in addition to examination and tests of summary statistics. The behavior of the winning margins during the 2012 and 2013 seasons, in comparison to those observed during the first five weeks of the 2014 season, is clearly depicted in Figure 2 below, constructed by means of Excel pivot tables. (See https://support.office.com/en-us/article/Create-a-PivotTable-and-analyze-your-data-7810597d-0837-41f7-9699-5911aa282760 for an introduction to Excel pivot tables.)

**Figure 2.** Average winnings margins by week, Sunday vs. Thursday games
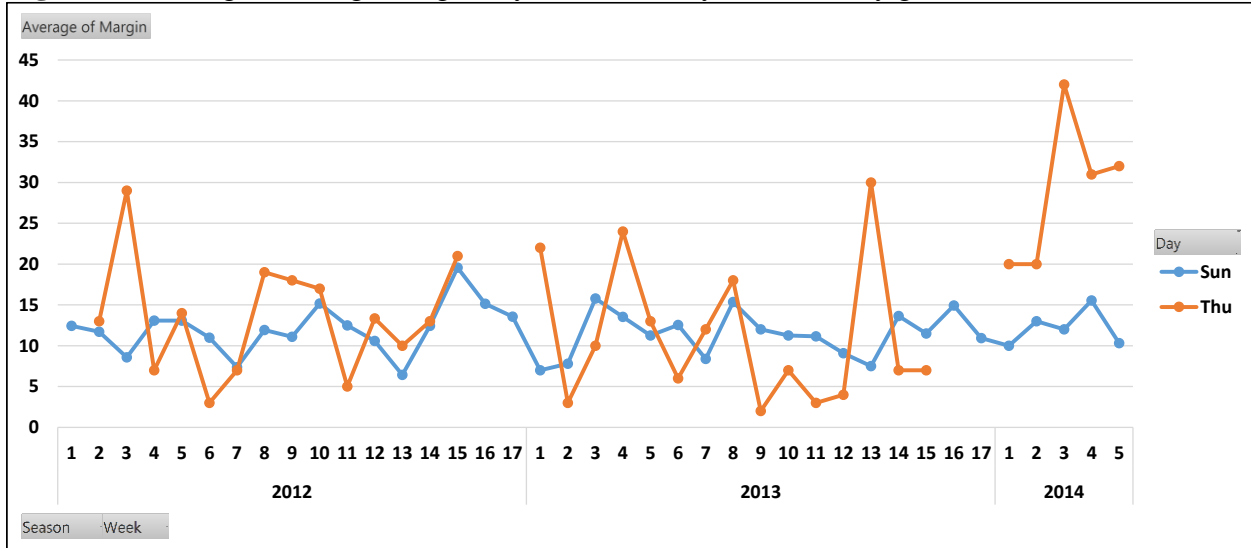
Here we see essentially stable random variation for both time series during the 2012 and 2013 seasons. The average Sunday margins obviously display less variability over time, as compared to the winning margin in the single Thursday game played each week. Nonetheless, we see the Thursday margins varying randomly above and below the Sunday average margin across the 2012 and 2013 seasons.

The first five weeks of the 2014 season, in contrast, display Thursday night margins consistently higher than the Sunday average margin. This is a visibly dramatic departure from the pattern observed over 2012 and 2013.

Indeed, even a simple non-parametric sign test suggests a statistically significant difference during the first five weeks of the 2014 season. Under the null hypothesis of no difference between the Sunday vs. Thursday game margins, we would expect a 0.5 probability each week that the Thursday game margin would be larger than the Sunday game margin, independent of the distributions of the respective nights' scores. Under this null hypothesis, the probability of seeing larger Thursday night margins five weeks in a row is $.5^5 = .03125$. (This test is, of course, less powerful than the *t* tests examined previously, but avoids the issue of the degree to which the normality assumption is met.)

This last observation prompts the question as to whether the Thursday night time series has experienced a significant change with the start of the 2014 season. Toward this end we construct a control chart, with 2σ and 3σ control limits based on the mean (12.39) and standard deviation (8.45) observed during the stable 2012 and 2013 seasons. The control chart appears in Figure 3 below.

Figure 3. Control chart for Thursday winning margins
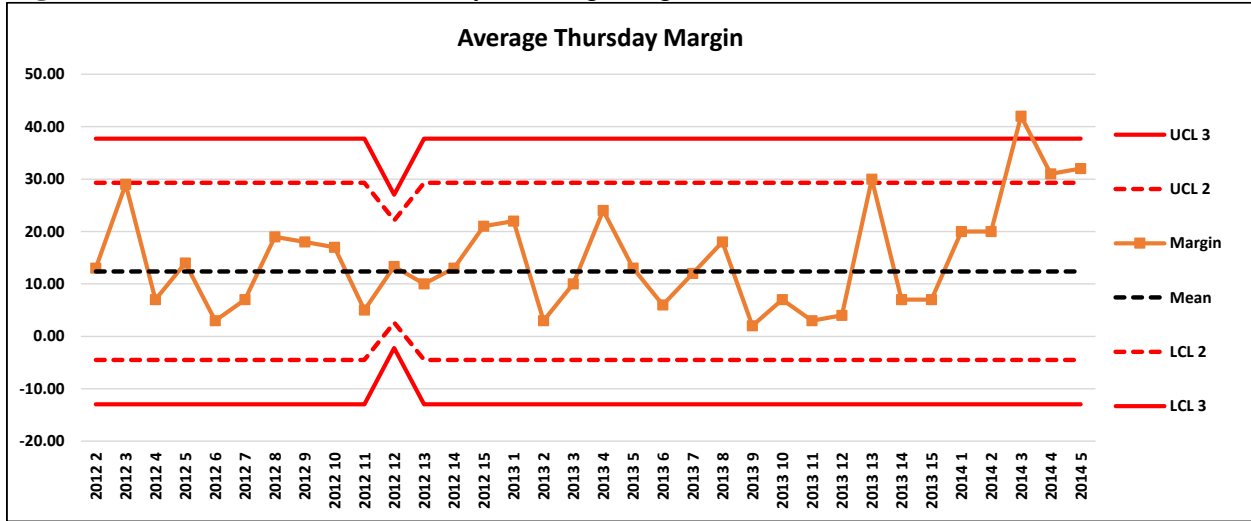
Note the 2012 and 2013 game margins appear to be "in control," with stable random variation, infrequently approaching the upper 2σ control limit due to the skewed nature of the data. (Week 12 of the 2012 season has narrower control limits as there were three Thursday games that week.) The 2014 season, in contrast, displays five points all above the mean, with one point above the 3σ upper control limit and three points in a row above the 2σ upper control limit. We thus have evidence that, not only are the 2014 season Thursday night games displaying larger winning margins, but this appears to be a new phenomenon unique to the 2014 season.

> *Teaching Hint: Be sure the students understand the point of the time series plots in Figure 2, vs. the control chart in Figure 3. Figure 2 is comparing the Sunday vs. Thursday night margins over time, and demonstrates that the Thursday night margins are consistently greater than the Sunday margins in 2014, while there was no such pattern in 2012-2013. Figure 3 is demonstrating that the sequence of Thursday night winning margins experienced a shift with the start of the 2014 season, e.g. there is something different happening that was not happening before. The control chart does not tell us what is different, it simply alerts us to the fact that "something" is different.*

4. Discussion

The preceding analysis leaves us with two possible scenarios: (a) there were some new mechanics influencing Thursday night game margins during the first five weeks of the 2014 season, distinct from those that governed Thursday night game margins during the 2012 and 2013 seasons, or (b) the first five weeks of the 2014 season saw an abnormal but nonetheless essentially random set of large Thursday night game margins, resulting in a "Type I Error" scenario in the hypothesis test.

As mentioned in the introduction to the paper, a great amount of speculation had been conducted in the popular media with regard to the causes of the large Thursday night margins. Given the preceding analysis, it is interesting to note that none of the explanations in the references provided earlier (shorter physical recovery time from Sunday to Thursday, shorter time to

develop game plan, etc.) could be argued to be unique to the 2014 season. Thus, proper analysis of historical data has changed the question from "Why are Thursday night game margins larger?" to "Why are Thursday night game margins larger <u>this year</u> when they have not been in previous years?" In essence, our understanding of common vs. special causes of variation has come into play.

> *Teaching Hint: Prior to pointing this out, again ask the student to speculate as to why the 2014 Thursday night game margins are larger than the Sunday game margins. After they have speculated for a while, ask them how many of those reasons could be considered <u>unique to the 2014 season</u>, and would not have been in effect during the 2012 and 2013 seasons.*

A famous quote is attributed to Nils Bohr, "Prediction is very difficult, especially if it's about the future." Given the timing of this writing, it will be interesting to see whether the pattern of large Thursday night margins continues through the 2014 season.

> *Teaching Hint: By the time this article is published, the 2014 NFL season will be complete, and the scores from the remainder of the season will be available at www.nfl.com/scores. It will be an interesting exercise for students to investigate whether the pattern of larger average Thursday (vs. Sunday) game margins persisted through the rest of the season. This exercise would also expose students to the process of data collection, whereas they typically start with data provided in a readily analyzed format.*

> *Recognize that if the larger Thursday margins have not persisted through the remainder of the season, the possibilities remain that (a) the pattern observed over the first five weeks of the 2014 season was a random fluke (generating Type I error in our hypothesis tests), or (b) there were factors in force during those five weeks leading to larger game margins, yet those factors did not persist through the entire season.*

> *Teaching Hint: Students will naturally attempt to extend the analysis in an attempt to build a model to predict who will win upcoming Thursday night games. Remind them that this was not the intent of the exercise, and we have very little data (only five observations in 2014) to support analysis of conditions that correlate with winning the Thursday night game.*

References

Carter, N., Felton, N., and Schwertman, N. (2014), "A Classroom Investigation of the Effect of Population Size and Income on Success in the London 2012 Olympics", *Journal of Statistics Education*, [online], 22, 2, available at http://www.amstat.org/publications/jse/v22n2/carter.pdf

Microsoft Corp (2014), "Create a Pivot Table and Analyze Your Data", [online], available at https://support.office.com/en-us/article/Create-a-PivotTable-and-analyze-your-data-7810597d-0837-41f7-9699-5911aa282760

NFL Enterprises LLC, (2014), "NFL Scores", [online], available at http://www.nfl.com/scores

Ruiz, S. (2014), "How to fix Thursday Night Football" [online], available at
http://q.usatoday.com/2014/10/03/how-to-fix-thursday-night-football-cbs-nfl-network/

Schalter, T. (2014), "What's to Blame for the NFL's Thursday Night Failures?" [online],
available at http://bleacherreport.com/articles/2219276-whats-to-blame-for-the-nfls-thursday-night-failures

Timothy S. Vaughan
University of Wisconsin – Eau Claire
College of Business
Eau Claire, WI 54702-4004
vaughats@uwec.edu