Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**BA 355: Business Analytics Case 1, Part 1**

Goal: To estimate the probability that the Denver Broncos beat the Seattle Seahawks on 9/8/24. More generally, estimate this probability for any game given the point spread.

The Broncos are currently a five point underdog (as of 8/25/24 at 9:33am) for the first game of the season, but the end result is by no means guaranteed. How would you estimate their chances, understanding that either team can win?

1. Combine that data for years 2013 – 2016 from the Excel file on the course webpage. You should have a total of 1068 rows. (A normal NFL season has 267 games including playoffs and the Super Bowl. 267\*4 = 1068)
2. Sort the data by point spread from largest (-26.5) to smallest (0).
3. Label column A as “Point Spread”, column B as “Favorite’s Score”, column B as “Underdog’s Score” and column D as “Over/Under” – we won’t need column D for now but might want it later.
4. Determine whether the favorite actually won the game (*not beat the spread*) for each point spread that occurred. In column E, subtract column C from column B. If this number is positive, it means the favored team won the game. If this number is negative, it means the underdog pulled off the upset. If this number is zero (and there are a couple of them), it means the game ended in a tie. Label this column as “Difference”.
5. In column, convert the result in column E to binary: a 1 if the favorite won the game and a 0 if the underdog won (or tied – for now we’re calling a tie a “win” for the underdog -- <https://en.wikipedia.org/wiki/1968_Yale_vs._Harvard_football_game>). Hint: =if. Label this column as “Win or Loss”. Or, maybe, it’s “better” to throw out the ties? Let’s discuss this in class and make a group decision. Let me know when you reach this stage.
6. Now, combine all the results for each possible point spread. For example, you should have 20 results from games with a 10-point spread; of these, the favorite won 16, but lost 4. What does this say about the *probability* that a 10-point favorite will win a game? Does 16/20 = 80% seem like a reasonable estimate? Combine the results so that for each possible point spread you know how many games there were (e.g. 20), how often the favorite won (e.g. 16) and the estimate for the probability (e.g. 80%). The best way to do this is by using the **PivotTable** function in MS Excel. If you don’t know how to do this, ask me or a classmate for help. If you do know how to do this, please share your expertise with your classmates.
7. In the end, you want a table of data with point spreads ranging from 0 to 26.5 in one column (but not every point is available, for example 26.5 is the only data point greater than 20) and the estimated probability of winning in the next column.

Now, use this data to make some estimates:

1. i) What proportion of 3 point favorites actually won?

ii) Is this a good estimate for the probability that a future 3 point favorite will win?

iii) What is the 95% confidence interval for the true proportion? [Extra Credit]

iv) Interpret this interval. [Extra Credit]

1. Repeat A. for 7 point favorites.
2. Repeat A. for 14 point favorites.
3. Does the trend make sense? Do higher favorites have a better chance of winning?
4. The Broncos are currently a five point underdog. According to your results, what percent of the time do teams favored by five win the game, like the Seahawks are in this game? So, since the Broncos are five point underdogs, what percent of the time do five point underdogs win?
5. Now, graph the point spreads (ranging from 0 to the max) versus the win probabilities and then fit a linear regression line to the data. Determine and interpret the equation of the line and the coefficient of determination (r2). **Print out** the graph with the line on it. How much does the probability of winning the game increase for every additional unit of point spread? At what point does the prediction from the line stop making sense?
6. Now fit a line from 0 to 12.5 points (up to where all probabilities become 100%) forcing the y-intercept to be 50%. (Why?). What is the equation of this line? Interpret both the y-intercept and the slope.
7. Use the equation of the line from G above to estimate the probability that 3 point, 7 point and 14 point favorites will win a game. Compare these to the empirical estimates from the data you found in parts A, B and C.
8. What is the probability for the favorite Seahawks? How does this compare to part E?
9. The graph is clearly **non-linear**. Use Excel to try to fit a better curve to the data. Soon we will develop our own function that should fit very well in class.
10. Repeat what you did in part F above, but now there are a lot more data points. For example, the point spread of 10 isn’t just one point with probability 80% it’s actually n = 20 points at point spread 10 with probability 80%. Use all 1068 data points. Calculate the coefficients of the new line to three decimal places – it will only be slightly different from what you got in part F.

**Deliverable:** Type up all of your results into a clear document that addresses each part above. The due date is TBD and will depend on the class’s progress and overall level of effort.

[Eventual Extra Credit: This data comes from the degenerate gambling website [www.goldsheet.com](http://www.goldsheet.com). This site stops recording data after the 2017 season. Find another reliable (for the internet) source that has data for the 2018, 2019 and 2020 seasons.]