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

1. i) What proportion of 3 point favorites actually won? **56%**

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

iii) What is the 95% confidence interval for the true proportion? (47%,64%)

iv) Interpret this interval. We are 95% sure real value is in this range.

1. Repeat A. for 7 point favorites. **72%, yes…**
2. Repeat A. for 14 point favorites. **100%, no…**
3. Does the trend make sense? Do higher favorites have a better chance of winning? **Yes but not monotonic due to random sampling error.**
4. The Broncos are currently a four point favorite. According to your results, what percent of the time do teams favored by four win the game? **61% but the Donkeys stink and lost!**
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). **The equation and r2 are on the graph. Starting at 54% for a pick ‘em game, the probability of winning increases 2.6% per unit of point spread. The r2 indicates that the point spread explains about ¾ of the probability of winning.** How much does the probability of winning the game increase for every additional unit of point spread? **2.6%** At what point does the prediction from the line stop making sense? **For point spreads > 17. No favorite is guaranteed to win.**
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?). **Why? We “know” it must start at 50/50, we are overriding what the data tells us.** What is the equation of this line? **y = -3%\*x + 50%** Interpret both the y-intercept and the slope. **Starting at 50%, prob of winning increases by 3% for every point of point spread.**
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. **59%, 71%, 92%. Close to data.**
8. What is the probability for the favored Broncos? **62%, close.** 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. **None of them fit.**
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.

**This is weighted. Can do multiple ways. Equation is now y = -0.0303\*x + 0.511.**