**Key**

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

1. i) 56%

ii) Yes. Or No with a good explanation.

iii) 56% ± 9%.

iv) We are 95% confident that if we collected a lot more data, the *real* proportion would be within the range 47% to 65%.

1. 72%. Yes, or No with a good explanation. 72% ± 12%. We are pretty certain the real probability that a touchdown favorite will win is between 60% and 84%
2. 100%. No, upsets happen. The confidence interval idea falls apart here.
3. Yes, there is a general trend for larger point spreads having higher probabilities of winning.
4. 37%.
5. The equation is y = 54% + 2.6%\*x with r2 = 77%. The y-intercept, 54% is weird since a 0 point spread game favorite should be at 50%. The slope, 2.6%, says that for every point of point spread, the favorite’s probability of winning increases by 2.6%. The r2 says that point spread explains 77% of the variability in probability of winning.

F.2 The equation is now y = 50% + 3%\*x after forcing the y-intercept to be 50%. Why? Because it makes sense, this model must start at 50% for an even game with no clear favorite should be 50/50. Now, starting from 50%, the probability of winning goes up 3% for every point of point spread.

F.3 You can do this in your head. 59%, 71% and 92%.

1. None of the pre-built Excel functions really work, but of course we know that the general logistic function will.
2. The equation is now y = 51% + 3%\*x, closer to what we expect for the y-intercept when we use all 1000+ data points. Same interpretation as before. The r2 value actually decreases to 66% (whereas it was 77% before). This somewhat ironic result (shouldn’t more data lead to a better fit???) happens when you add more data which is why some statisticians recommend using the “Adjusted r-squared” instead of just “r-squared.” Google this if you’re interested.