Problem Set 1

Donovan Doyle

1/30/2019

Question 1

A) The minimum distance was 18. The maximum was 76, but it missed. The mean was 36.90. The median was 37. **B)** The minimum can't get lower than 18 because for each kick, you must also include the distance of the end zone and the distance between the longsnapper and the holder. The distance of the end zone is 10 yards, and the holder will generally set up 7 yards behind the longsnapper, so any kick, even from the 1 yard line, will be at least 18 yards. The maximum can be explained because it occurred in the 30th minute, meaning right at halftime. The Raiders probably thought they had a better chance trying a kick than they did throwing a deep ball to the endzone.

Question 2

The percentage of kicks made between 40 and 45 yards was 79.2%, while the percentage of kicks made above 45 was 64.4%.

Question 3

The make rate on grass was 82%, while the make rate on turf was slightly higher at 84%. The difference is statistically significant at a 99% level when tested with no control variables. Even with control variables added (I used GameMinute, ScoreDiff, and Distance), the relationship is still statistically significant, but it is much smaller than the 2% in the raw make rates. 2% is not the true effect of surface, but there is an effect.

Question 4

A) The correlation between grass and distance is -0.003. This likely means when there's turf, a coach is very slightly more willing to take a kick. **B)** The correlation between success and distance is -0.337. This means the longer a kick, the less likely it is to be made.

Question 5

A) Covariance of X and Y over the variance of X is the formula for OVB. **B)** Distance is confirmed OVB, as it is independent from X (grass) and impacts Y (success). Because it has a negative correlation with grass, teams are less likely to kick long distances if they're on a grass field. When running a regression with the dependent variable being "Success", the independent variable being "Grass" and the control variable being "Distance", there's a negative coefficient on Distance, and the negative coefficient on grass becomes larger, meaning that when distance is controlled for, the true effect of the surface increases.

Question 6

A) The Game Minute is not statistically significant compared to the success rate, so there seems to be no significant correlation between the two. There is no evidence of "clutch" kicking. **B)** This corrects for skill of the kicker, isolating the effects of the Game Minute on the success rate. The adjusted R-squared is now 0.1205, greater than the old adjusted R-squared of 0.1139. **C)** It seems kickers have been getting better over the years, as 2012-2015 all have positive significant relationships with the make rate. This is in line with most other athletic skills, as kickers get better over time, such as a 4-minute mile supposedly being impossible or the average weight of an offensive lineman increasing over the years.

Question 7

A) There is an 88.87% chance that Tucker makes the field goal, given the conditions. **B)** Yes, this makes sense. I would assume it would be slightly under his average, as the ScoreDiff and GameMinute would skew his make rate down, given that you generally have to take longer chances at GameMinute 30 and in that close of a game, meaning he would have to kick longer field goals in that situation usually.

Question 8

A) I got 88.87% chance again. **B)** I think this may be because R auto-runs a logistic when the dependent variable is between 0 to 1.

Question 9

Clustered standard errors are used when standard errors are correlated to each other in panel data. For example, in our case, a kicker's standard error from one year to the next will be related to each other, because if he is kicking in 2015 after playing in 2014, he likely had a pretty good year in 2015, or at least a good enough year to not be benched. For this reason, I would cluster at the kicker factor level, as that will likely be correlated with each other more than at the year factor level.

Question 10

Kicker	Make Rate	Kick Count
Akers	0.81	336
Brown	0.84	488
Bryant	0.87	308
Crosby	0.83	321
Dawson	0.88	332
Gostkowski	0.89	342
Gould	0.85	329
Janikowski	0.82	334
Vinatieri	0.87	339

I took the kickers with the 9 most kicks in the time period, as volume is important in my opinion, showing that the coach trusts them and they never were beaten out. Of these kickers, Gostkowski has the highest make rate. This provides solid evidence that Gostkowski is the best kicker of the time period, although the

numbers are very close. I think Brown would also be a good choice, given his workload, his coaches must trust him. In the regression that includes clustering kickers, which contains controls for all of our quantifiable variables, Tucker and Vinatieri have the strongest relationship between their kicking and success.

Question 11

I would want 1) weather results, 2) playoff versus regular season results, and 3) blocked kicks/messy snaps and holds vs clean kicks. Controlling for weather results would positively affect the make rates, as bad weather is bad for kicking, playoff results would help see the "clutch" gene better if it exists, and controlling for only clean kicks would positively affect the make rates.

I used this data from Tony Crabtree that has weather by game for another project, which would be helpful if I wanted to dig more:

https://www.kaggle.com/tobycrabtree/nfl-scores-and-betting-data

All of my work in R is contained here: https://github.com/donovandoyle/1042_PS1