Tristan Conde

ADTA 5160

Spring 2022

# Assignment 02

**Part 1:**

* Examine the raw data and the variable descriptions. Produce summary statistics for the following variables: *fired, winpcttenure, gamescoached, ageathire,* *yrsnflhcexpathire, playedinnfl, offcoach, black*

Table

Description automatically generated

* Briefly describe findings that you think are relevant from this summary statistics table.
  + The average coach in our dataset coached about 68 games during their employment spell.
  + The average age of coaches during their time of hire was quite higher than expected at around 50 years of age.
  + Fired has a significantly high mean around 0.72, which would imply that a significant number of coaches within this dataset have been fired or let go in some capacity.
  + Higher mean average for the offcoach variable lets us know that there was a decent amount of coaches have some offensive experience
  + Lower mean in the playedinnfl variable says that the dataset has a lot of coaches that did not play themselves within the NFL.
* What percentage of employment spells in the sample come from *white* head coaches*?* How do you know this? Explain.

Table

Description automatically generated

* + Through a tabulate command of the nonblack variable, we are able to determine that there are 190 coaches that are non-black, which equates to 87.96% (88%)
  + To get more granular, that can be done via the tabulate white command in Stata, which gives us that 86.57% (87%) of coaches in our dataset are white.
* Create an interaction term between the variable that identifies whether or not the HC played quarterback as a player and the variable that represents the total number of years of NFL coaching experience for the HC at the time of hire.

Table

Description automatically generated

* Assume you are interested in identifying the determinants of whether or not a HC is fired to conclude his employment spell. The limitation is that you are only able to include six right-hand side predictor variables. Make sure you choose the six that you believe are the most appropriate in their expected relationship with a HC being fired.
  + First, decide what regression-based modeling approach is most appropriate to answer this question.
    - I decided to use the Logit regression model for this analysis.
  + Estimate your selected model.

Table

Description automatically generated

Table

Description automatically generated

* + In three sentences or less, explain why you selected this specific empirical approach.
    - Logit regression is the approach I decided to take since the dependent variable is a binary choice of 1 (meaning they got fired) or 0 (not fired) at the end of their employment spell. As such, a logit model is appropriate since we are looking for if these variables in a way are probable causes for the event happening.
  + Then explain why you selected each of your six independent variables (*in two sentences or less for each variabl*e).
    - Winpcttenure – I selected this variable because I believe that if a coach has had a lower win percentage over their employment spell, that would incite management to fire the head coach.
    - Emptenureseasons – I selected this variable as I believe that the amount of tenure or time that a coach has been with a company also does play a factor. It is a possibility that instead of firing a coach with long tenure they may move him to another position or allow him to retire instead of firing.
    - Atsfinalyr – I had selected this variable based on this would show their ats during the final year, and just like with winpcttenure, if this was lower as well, it might have been causation for firing.
    - Allnflexp – I selected this variable on the thought process that management may or may not fire a head coach after thinking they just may not have enough experience and thus want to grab another HC that has more.
    - Teamwpcthire5 – I selected this variable based on believing that there is a possibility head coaches can and will be fired if the team stats over the past 5 years seem to be an improvement or higher than the current stats brought on by the current head coach’s employment spell.
    - Ageathire – I selected this variable since I went under the process of assuming that there are some instances where head coaches may be fired or let go due to their age, which itself may bring its own set of problems; being health, mental, logical, etc.
* How many observations were included in the estimation of this model based on the model statistics? Is this number different than the number of observations in the Excel file which includes the raw data? Explain why or why not.
  + Our model has 210 observations versus the 216 that is within the raw data of Excel. This is due to some of our variables (teamwpcthire5) having missing or incomplete data, and thus is left out of the model.
* For each independent variable included in the regression, in two sentences or less, explain the result of each coefficient as if you were writing up the results in a report to a stakeholder. Your two sentences should tell the reader whether or not the independent variable is statistically significant and should explain the relationship between the independent variable and the dependent variable.
  + Winpcttenure – This variable is statistically significant at the 1% level, as it has a p-value below 0.0005, and has an effect size of -8.26 and a dy/dx of -1.27. This tells us that this variable decreases the fired variable by -1.2733.
  + Emptenureseasons – This variable ends up with a very high p-value of 0.623, therefore leading us to infer that it is not statistically significant. This variable has an effect size of 0.0287 and a dy/dx of .0044202, which tells us that this variable increases the fired variable by 0.0044202.
  + Atsfinalyr – This variable concludes a p-value of 0.017, which is statistically significant at the 5% level. The variable has an effect size of -3.91 and a dy/dx of -.6026374, which tells us that this variable decreases the fired variable by -0.6026374.
  + Allnflexp – This variable has a p-value of 0.119, which signifies that this variable is not statistically significant. The variable also has an effect size of 0.499 and a dy/dx of 0.077, meaning that this variable increases the fired variable by 0.0077061.
  + Teamwpcthire5 – This variable has a p-value of 0.208, which gives away that this variable is not statistically significant. This variable also has an effect size of 1.978 and a dy/dx of 0.3048, implying that this variable increases the fired variable by 0.304776.
  + Ageathire – This variable has a p-value of 0.097, which is statistically significant at the 10% level. This variable also has an effect size of -0.480 and a dy/dx of -0.0074, meaning that this variable decreases the fired variable by -0.0074068.
* Assume you are able to estimate an updated model that includes eight right-hand side predictors instead of six.
  + Which two additional independent variables would you include? Explain your reasoning in four sentences or less.
    - Gamescoached – Adding gamescoached I believe is another predictor that I think would help add to the regression model. The reason that I believe this variable would be helpful as I believe the amount of games that a head coach has coached would be something that management and the team would look at. Looking at that along with the winning percentage and ATS record of the team during their employment spell could align and point to them firing said coach is the performance and stats are not to their liking.
    - Teamtenure – Teamtenure is a variable that I would have also liked to include. For the reasoning, it is very similar to the reasons that emptenureseasons is in the regression model to begin with, as I believe management may be a bit more lenient and go other ways to let a head coach go depending on how long they have been a HC with them or just with the organization.

**Part 2:**

* Assume you are interested in identifying the determinants of team performance during NFL HC employment spells. Again, you are only able to include six right-hand side predictor variables.
  + First, decide what modeling approach is most appropriate to answer this question. Make sure you choose the six that you believe are the most appropriate in their expected relationship with team performance during HC employment spells.
    - I decided to go with the Fractional Regression model for this analysis.
  + Estimate your selected model.

Table

Description automatically generated

* + In three sentences or less, explain why you selected this specific empirical approach.
    - The reason I chose this specific approach is based on the dependent variable I chose, which is the winning percentage of the HC during their employment spell. This variable showcases the percentage of wins that the head coach has during their employment tenure. Since this percentage is calculated through a 0 (losing the game) and a 1 (winning the game), fractional regression is the appropriate model.
  + Then explain why you selected each of your six independent variables (in two sentences or less for each variable).
    - Emptenureseasons – I selected this variable because I believe that with an head coach that is with a team for a long time will build camaraderie with their players and will be able to build on their performance by knowing what players strengths and weaknesses are.
    - Atsrecord – I selected this variable since the ATS record is made up of the win/loss record in the games where the point spread was available. Since this has a lot to do with win percentage, I thought this variable was very appropriate.
    - Gamescoached – I selected this variable due to the thought that having more games coached by the head coach, the more experience they will have in the NFL scene and with the players. Thusly, this will likely help win percentage and performance since they will have more time in the game itself.
    - Teamtenure – I selected this variable due to similar reasons that I picked the emptenureseasons variable. Having longer amounts of times with the team will help the head coach to work around teammates strengths and weaknesses and can lead to better performance.
    - Incomeraw – I selected this variable due to the income of the market within the team would be a significant variable in terms of what players and budget the team can afford to get. It can very likely make or break a team since teams with higher budgets can afford more star-studded players than others.
    - Teamwpcthire5 – I selected this variable as the performance of the players matter just as much when it comes to the head coach’s performance. The team performance and winning percentage over the past five years comes from a mixture of not only coaching, but from the players as well, so it should be taken into account when trying to determine performance.
* How many observations were included in the estimation of this model according to the model statistics? Is this number different than the number of observations in the Excel file which includes the raw data? Explain why or why not.
  + Our model has 204 observations versus the 216 that is within the raw data of Excel. This is due to some of our variables (teamwpcthire5 & incomeraw) having missing or incomplete data, and thus is left out of the model.
* For each independent variable included in the model, in two sentences or less, explain the result of each coefficient as if you were writing up the results in a report to a stakeholder. Your two sentences should tell the reader whether or not the independent variable is statistically significant and should explain the relationship between the independent variable and the dependent variable.
  + - Emptenureseasons – This variable has a lower p-value of 0.023 and it is statistically significant at the 5% level. The variable also has a coefficient of 0.5439 and a dy/dx of 0.1271, meaning that as this variable increases, it increases the winpcttenure variable by 0.1271.
    - Atsrecord – This variable has a lower p-value of below 0.00005, which signifies that this variable is statistically significant at the 1% level. The variable also has a coefficient of 5.5354 and a dy/dx of 1.2935, meaning that as this variable increases, it increases the winpcttenure variable by 1.2935.
    - Gamescoached – This variable has a p-value of 0.042, signifying that it is statistically significant at the 5% level. The variable also has a coefficient of -0.03056 and a dy/dx of -0.00714, meaning that as this variable increases, it decreases the winpcttenure variable by -0.00714.
    - Teamtenure – This variable has a very high p-value of 0.267, which signifies that it is not statistically significant for this analysis. The variable also has a coefficient of -0.00180 and a dy/dx of -0.000420, meaning that as this variable increases, it decreases the winpcttenure variable by -0.000420.
    - Incomeraw – This variable has very high p-value of 0.307 which signifies that it is not statistically significant for this analysis. The variable is also extremely interesting, as it has a coefficient of 2.36e-06 and a dy/dx of 5.52e-07. These variable numbers are so small, it infers that this variable affects the dependent variable so mildly and therefore as it increases it has little to no effect.
    - Teamwpcthire5 – This variable has a lower p-value of below 0.00005, which signifies that this variable is statistically significant at the 1% level. The variable also has a coefficient of 1.3061 and a dy/dx of 0.30520, meaning that as this variable increases, it increases the winpcttenure variable by 0.30520.
* Assume you are able to estimate an updated model that includes eight right-hand side predictors instead of six. Which two additional independent variables would you include? Explain your reasoning in four sentences or less.
  + Playedinnfl – I would add this variable because I think that having the NFL playing experience would be beneficial if a player made the transition to head coach. Having the experience of being a player on the field I think could help translate to making sure that your players understand everything you’re planning and thinking and also maybe a few extra tricks or plays those other coaches can’t think of since you were on the field and can use that to their advantage.
  + Hcwpctathire – I would add this variable since I believe this would showcase possibly why a head coach would be hired for a new team in the first place. A higher winning percentage could also provide additional insights for our model implying that a head coach with a higher winning percentage could be a significant model for winning percentage of any team that they are a part of.