Foundations of Sport Analytics: Data, Representation, and Models in Sports.

Weekly Quizzes

Week 1 — Quiz 1

Question 1

How many EPL games from this season were played in 2018? (1 point)

a. 190

b. 171

c. 380

d. 209

Question 2

Which team scored the highest number of goals while playing at home in the first half of the season? (1 point)

a. Manchester City

b. Liverpool

c. Watford

d. Stoke

Question 3

Which team conceded the highest number of goals while playing away in the first half of the season? (1 point)

a. Liverpool

b. Watford

c. Stoke

d. West Ham

Question 4

Which of the following teams had the smallest difference between their win percentage and Pythagorean expectation in the first half of the season? (1 point)

a. Manchester City

b. Manchester United

c. Arsenal

d. Liverpool

Question 5

Which of the following teams had the smallest difference between their win percentage and Pythagorean expectation in the [first] half of the season? (1 point)

a. Bournemouth

b. Stoke

c. Leicester

d. Brighton

Question 6

Which of the following teams had the highest value for away wins (awinvalue) for in the first half of the season? (1 point)

a. West Ham

b. West Brom

c. Stoke

d. Crystal Palace

Question 7

Which team had the largest gap between home points won (hwinvalue) and away points won (awinvalue) in the second half the season? (1 point)

a. West Ham

b. Brighton

c. Watford

d. Arsenal

Question 8

What was the correlation between win percentage and the Pythagorean expectation in the first half of the season? (1 point)

a. 0.956

b. 0.968

c. 1.000

d. 0.796

Question 9

What was the correlation between win percentage in the first half of the season and the second half of the season? (1 point)

a. 0.968

b. 0.757

c. 1.000

d. 0.796

Question 10

What was the correlation between win percentage in the second half of the season and the Pythagorean expectation in the first half of the season? (1 point)

a. 1.000

b. .796

c. 0.746

d. .757

Week 2 — Quiz 1

Question 1

What are the number of observations and the number of variables in the NHL_Game dataframe after performing the first 7 steps? (1 point)

a) 18,946; 13

b) 18,946; 24

c) 18506; 24

d) 18506; 13

Question 2

What is the time range of the NHL_Game dataframe after you performed step 8? (1 point)

a) 2015-03-08 to 2018-06-08

b) 2010-10-07 to 2018-06-08

c) 2010-10-07 to 2018-06-14

d) 2010-10-07 to 2015-03-08

Question 3

After performing step 9 above, what are the values of the "gid" variable of the fifth, tenth, and fifteenth observations by date in ascending order in the prepared NHL_Game dataframe? (1 point)

a) 5666, 5662, 5668

b) 2730, 2725, 2720

c) 5662, 5668, 5683

d) 2725, 2720, 2716

Week 2 — Quiz 2

Question 1

What are the mean and standard deviation of the total number of goals for in the "NHL_Game" dataframe? (1 point)

a) 2.962; 1.689

b) 2.721; 0.466

c) 2.826; 1.655

d) 2.690; 1.609

Question 2

What is the mean of the total number of goals against for home games? What is the mean of the total number of goals against for away games? (1 point)

a) Home game: 2.690; Away game: 2.962

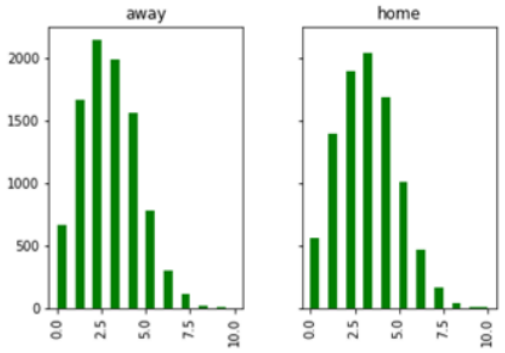
b) Home game: 2.841; Away game: 2.642

c) Home game: 2.721; Away game: 2.828

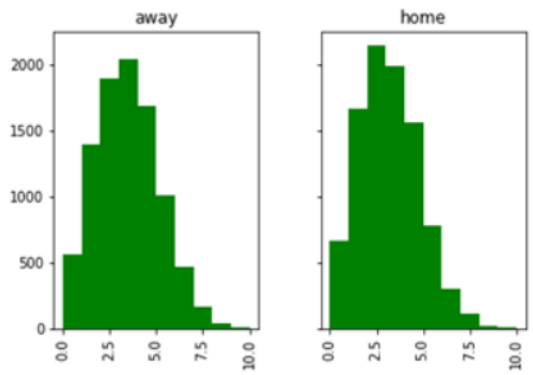
d) Home game: 2.840; Away game: 2.840

Question 3

Which of the following histograms is the correct graph for step two? (1 point)

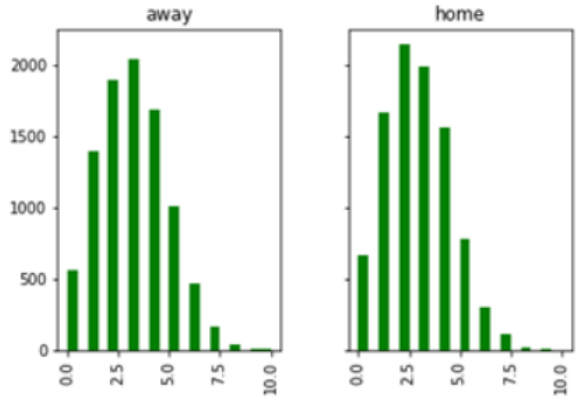


Two histograms that show the distribution of the number of goals against, one for away games, one for home games. For both histograms, the horizontal axis represents the value of the number of goals against, ranging from 0 to 10; the vertical axis represents the number of games having the number of goals against in a given range. The distribution of the goals against in both graphs is slightly skewed right. For away games, the number of goals against centers and peaks at about 2. For home games, the number of goals against centers and peaks at around 3. For both graphs, there are spaces between bins.

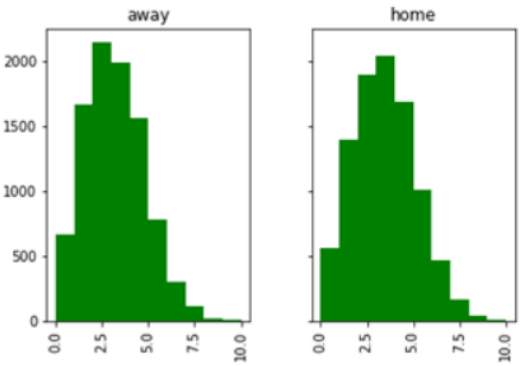


Two histograms that show the distribution of the number of goals against, one for away games, one for home games. For both histograms, the horizontal axis represents the value of the number of goals against, ranging from 0 to 10; the vertical axis represents the number of games having the number of goals against in a given range. The distribution of the goals against in both graphs is slightly skewed right. For away games, the number of goals against centers and

peaks at about 3. For home games, the number of goals against centers and peaks at around 2. For both graphs, there is no space between bins.

◉ 

Two histograms that show the distribution of the number of goals against, one for away games, one for home games. For both histograms, the horizontal axis represents the value of the number of goals against, ranging from 0 to 10; the vertical axis represents the number of games having the number of goals against in a given range. The distribution of the goals against in both graphs is slightly skewed right. For away games, the number of goals against centers and peaks at about 3. For home games, the number of goals against centers and peaks at around 2. For both graphs, there are spaces between bins.

○ 

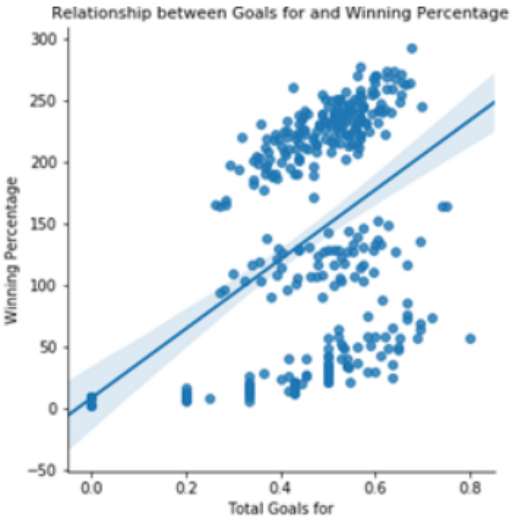
Two histograms that show the distribution of the number of goals against, one for away games, one for home games. For both histograms, the horizontal axis represents the value of the number of goals against, ranging from 0 to 10; the vertical axis represents the number of games having the number of goals against in a given range. The distribution of the goals against in both graphs is slightly skewed right. For away games, the number of goals against centers and peaks at about 2. For home games, the number of goals against centers and peaks at around 3. For both graphs, there is no space between bins.**

Week 2 — Quiz 3

Question 1

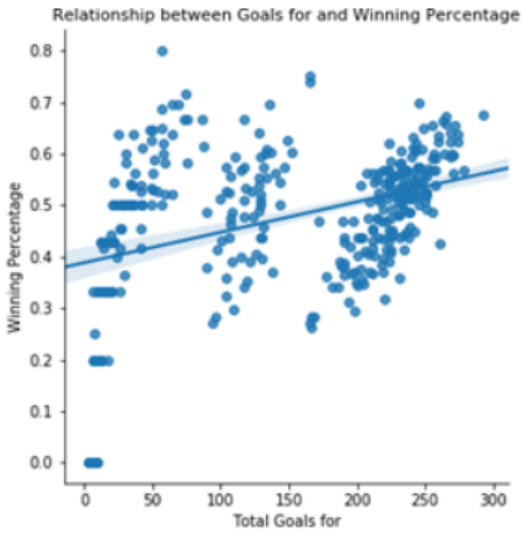
Which of the following graph is correct for step 1? (1 point)

This is a scatterplot with an approximation line. The horizontal axis represents the total *goals for* and the vertical axis shows winning percentage. There are two clusters of points. The horizontal axis of both clusters of points ranges from approximately 0.25 to 0.8. For the vertical axis, one cluster ranges from approximately 90 to 150 and other cluster ranges from around 160 to 250. There are more points in the second cluster than the first one. For both clusters, the points scatter in a pattern beginning from the lower left corner with a slight positive slope. This suggest that within each cluster, the larger the number of goals for, the higher the winning percentage. The approximation line has a vertical intercept of approximately 160 and the value of y-axis increases to around 250 when the value of the x axis is 0.8.



This is a scatterplot with an approximation line. The horizontal axis represents the total *goals for* and the vertical axis shows winning percentage. There are three clusters of points. The horizontal axis of the first cluster ranges from 0 to 0.8 while the horizontal axis of the other two clusters ranges from approximately 0.2 to 0.75. For the vertical axis, the first cluster ranges from approximately 0 to 70, the second cluster

ranges from approximately 70 to 150, and the last cluster ranges from around 150 to 270. There are more points in the third cluster than the first two. For all three clusters, the points scatter in a pattern beginning from the lower left corner with a slight positive slope. This suggest that within each cluster, the larger the number of goals for, the higher the winning percentage. The approximation line has a vertical intercept of approximately 0 and the value of y-axis increases to around 250 when the value of the x-axis is 0.8.

○ 

This is a scatterplot with an approximation line. The horizontal axis represents the total *goals for* and the vertical axis shows winning percentage. There are three clusters of points. The vertical axis of the first cluster ranges from 0 to 0.8 while the vertical axis of the other two clusters ranges from approximately 0.25 to 0.75. For the horizontal axis, the first cluster ranges from approximately 0 to 70, the second cluster ranges from approximately 70 to 150, and the last cluster ranges from around 150 to 270. There are more points in the third cluster than the first two. For all three clusters, the points scatter in a pattern beginning from the lower left corner with a steep positive slope. This suggest that within each cluster, the larger the number of goals for, the higher the winning percentage. The approximation line has a vertical intercept of approximately 0.39 and the value of y-axis increases to around 0.55 when the value of the x-axis is 300.

This is a scatterplot with an approximation line. The horizontal axis represents the total *goals for* and the vertical axis shows winning percentage. There are two clusters of points. The vertical axis of both clusters of points ranges from approximately 0.25 to 0.8. For the horizontal axis, one cluster ranges from approximately 90 to 150 and other cluster ranges from around 160 to 250. There are more points in the second cluster than the first one. For both clusters, the points scatter in a pattern beginning from the lower left corner with a steep positive slope. This suggest that within each cluster, the larger the number of goals for, the higher the winning percentage. The approximation line has a vertical intercept of approximately 0.43 and the value of y-axis increases to around 0.55 when the value of the x-axis is 300.

Question 2

What is the correlation coefficient between total number of goals for and winning percentage from step 2? (1 point)

a) 0.749

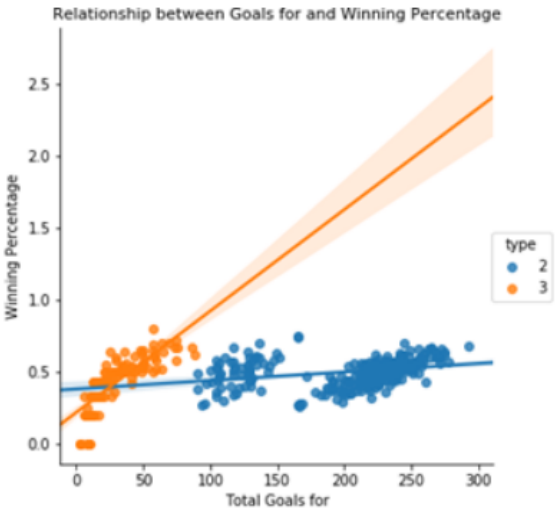
b) 0.408

c) 0.316

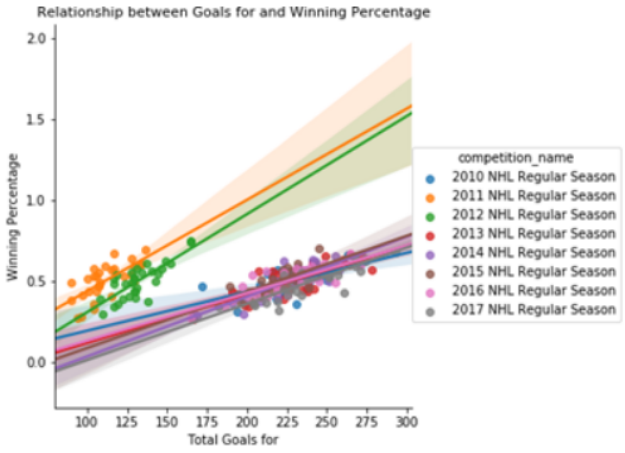
d) 0.619

Question 3

Which of the following is the correct graph for part 3? (1 point)

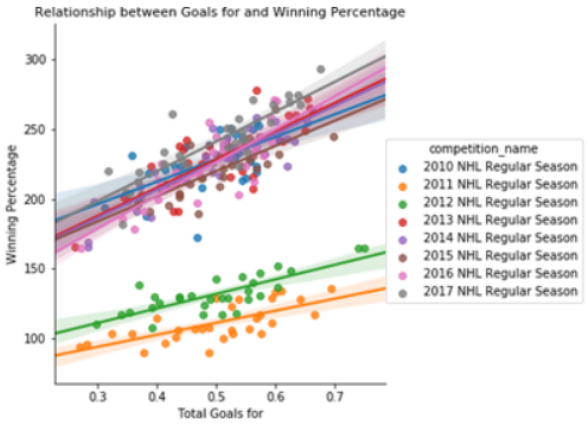


This is a scatterplot with approximation lines. The horizontal axis represents the total *goals for* and the vertical axis shows winning percentage. The points are grouped by the types of the games. Each type is represented by a different color. For points of type 3 games, the vertical axis ranges from approximately 0 to 0.8 and the horizontal axis ranges between approximately 0 and 90. The approximation line intercepts the y-axis at around 0.1 and the value of the y-axis increases to about 2.4 when the value of the x-axis is 300. For points of type 2 games, there are two clusters. The vertical axis of both clusters of points ranges from approximately 0.25 to 0.8. For the horizontal axis, one cluster ranges from approximately 90 to 150 and other cluster ranges from around 160 to around 290. There are more points in the second cluster than the first one. The approximation line has a vertical intercept of approximately 0.4 and the value of y-axis increases to around 0.55 when the value of the x-axis is 300. The slope of the approximation line for type 3 games is much steeper than the slope of the approximation line for type 2 games.



This is a scatterplot with approximation lines. The horizontal axis represents the total *goals for* and the vertical axis shows winning percentage. The points are grouped by the names of the competition. Each competition is represented by a different color. The vertical axis ranges from approximately 0.25 to 0.8. The points for 2011 and 2012 regular seasons have smaller number of *goals for* and cluster to the left of the points for all other seasons. The horizontal axis for the 2011 and 2012 season ranges between around 90 and 175. The horizontal axis for the other season ranges between 160 and 275. The approximation lines for 2011 and 2012 regular

seasons have slightly higher vertical intercepts, between 0.25 and 0.4, and slightly steeper slopes. For the other seasons, the approximation lines intercept the y-axis approximately between -0.1 and 0.15 and the values of the y-axis increase to around 0.5 when the value of the x-axis is 300.

○ 

This is a scatterplot with approximation lines. The horizontal axis represents the total *goals for* and the vertical axis shows winning percentage. The points are grouped by the names of the competition. Each competition is represented by a different color. The horizontal axis ranges from approximately 0.25 to 0.8. The points for 2011 and 2012 regular seasons have lower winning percentage and cluster to below the points for all other seasons. The vertical axis for the 2011 and 2012 season ranges between around 90 and 175. The vertical axis for the other season ranges between 160 and 275. The approximation lines for 2011 and 2012 regular seasons have lower vertical intercepts, between 90 and 105, and slightly smaller slopes. For the other seasons, the approximation lines intercept the y-axis approximately between 160 and 180 and the values of the y-axis increase to between 250 and 300 when the value of the x-axis is 0.8.

○ 

This is a scatterplot with approximation lines. The horizontal axis represents the total *goals for* and the vertical axis shows winning percentage. The points are grouped by the types of the games. Each type is represented by a different color. For points of type 3 games, the vertical axis ranges from approximately 0 to 90 and the horizontal axis ranges between approximately 0 and 0.8. The approximation line intercepts the y-axis at around -10 and the value of the y-axis increases to about 50 when the value of the x-axis is 0.8. For points of type 2 games, there are two clusters. The horizontal axis of both clusters of points ranges from approximately 0.25 to 0.8. For the vertical axis, one cluster ranges from approximately 90 to 150 and other cluster ranges from around 160 to around 290. There are more points in the second cluster than the first one. The approximation line has a vertical intercept of approximately 105 and the

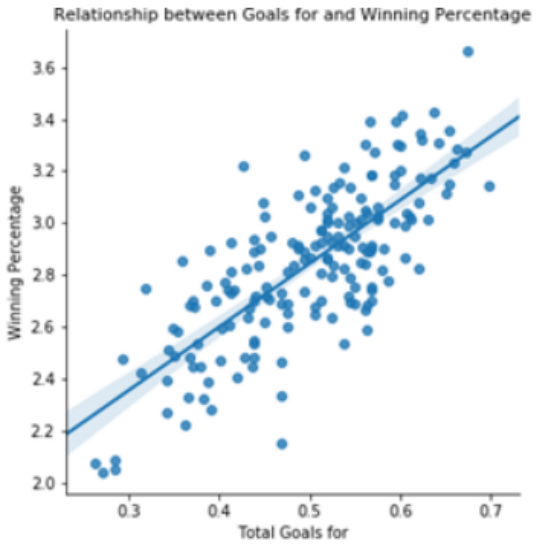
value of y-axis increases to around 250 when the value of the x-axis is 0.8. The slope of the approximation line for type 2 games is slightly steeper than the slope of the approximation line for type 3 games.

Question 4

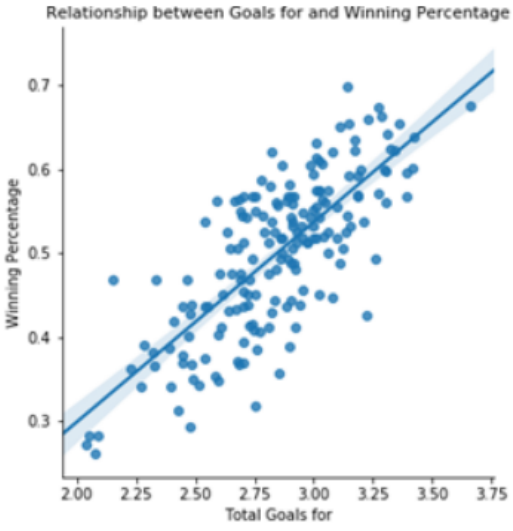
Which of the following is the correct graph for part 5? (1 point)

○ 

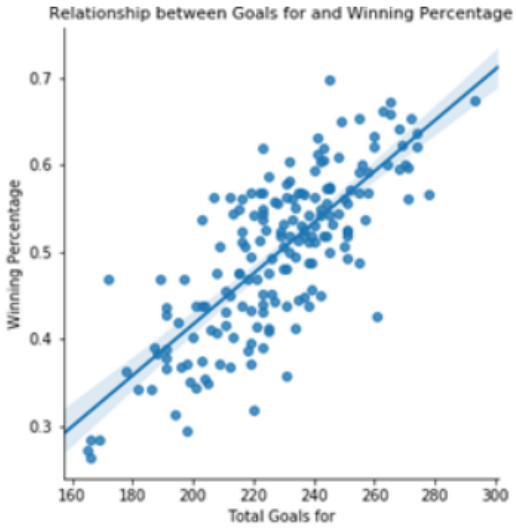
This is a scatterplot with an approximation line. The horizontal axis represents the total *goals for* and the vertical axis shows winning percentage. The horizontal axis shows total goals for ranging between approximately 0.25 and 0.8. The vertical axis shows winning percentage shows winning percentage ranging between approximately 165 and 300. The points scatter in a pattern beginning from the lower left corner with a positive slope. This suggest that the larger the number of goals for, the higher the winning percentage. The approximation line has a vertical intercept of approximately 175 and the value of y-axis increases to around 275 when the value of the x-axis is 0.8.

○ 

This is a scatterplot with an approximation line. The horizontal axis represents the total *goals for* and the vertical axis shows winning percentage. The horizontal axis shows total goals for ranging between approximately 0.25 and 0.8. The vertical axis shows winning percentage shows winning percentage ranging between approximately 2 and 3.65. The points scatter in a pattern beginning from the lower left corner with a positive slope. This suggest that the larger the number of goals for, the higher the winning percentage. The approximation line has a vertical intercept of approximately 2.2 and the value of y-axis increases to around 3.4 when the value of the x-axis is 0.8.



This is a scatterplot with an approximation line. The horizontal axis represents the total *goals for* and the vertical axis shows winning percentage. The horizontal axis shows total goals for ranging between approximately 2 and 3.65. The vertical axis shows winning percentage shows winning percentage ranging between approximately 0.25 to 0.8. The points scatter in a pattern beginning from the lower left corner with a positive slope. This suggest that the larger the number of goals for, the higher the winning percentage. The approximation line has a vertical intercept of approximately 0.3 and the value of y-axis increases to around 0.8 when the value of the x-axis is 3.75.



This is a scatterplot with an approximation line. The horizontal axis represents the total *goals for* and the vertical axis shows winning percentage. The horizontal axis shows total goals for ranging between approximately 165 and 300. The vertical axis shows winning percentage shows winning percentage ranging between approximately 0.25 to 0.8. The points scatter in a pattern beginning from the lower left corner with a positive slope. This suggest that the larger the number of goals for, the higher the winning percentage. The approximation line has a vertical intercept of approximately 0.3 and the value of y-axis increases to around 0.8 when the value of the x-axis is 300.

Question 5

What is the correlation coefficient between total number of goals for and winning percentage in part 6? (1 point)

a) 0.776

b) 0.761

c) 0.753

d) 0.771

Week 3 — Quiz 1

Question 1

Which of the four players had the greatest number of shots in the season? (1 point)
a) Dwight Howard
b) Russell Westbrook
c) DeAndre Jordan
d) Kevin Durant

Question 2

Comparing the plots of Russell Westbrook and Kevin Durant, which of the following statements is best supported by the images? (1 point)
a) Kevin Durant takes more three point shots that Russell Westbrook
b) Russell Westbrook has a higher success rate than Kevin Durant
c) Kevin Durant has a higher success rate Russell Westbrook
d) Russell Westbrook shoots more than Kevin Durant

Question 3

Comparing the plots of DeAndre Jordan and Dwight Howard, which of the following statements is best supported by the images? (1 point)
a) DeAndre Jordan takes more shots at a distance from the basket than Dwight Howard
b) Dwight Howard takes more shots at a distance from the basket than DeAndre Jordan
c) DeAndre Jordan is a better shooter than Dwight Howard
d) Dwight Howard is a better shooter than DeAndre Jordan

Question 4

Comparing the plots of Brook Lopez and Brian Lopez, which of the following statements is best supported by the images? (1 point)
a) The Lopez twins never attempt 3 point shots
b) Robin Lopez attempts 3-point shots but Brook Lopez doesn't
c) Brook Lopez attempts 3-point shots but Robin Lopez doesn't
d) Both of the Lopez twins attempt 3 point shots

Question 5

Based on these plots, which of the six players seems most likely to found shooting from the left hand side of the basket rather than the right hand side? (1 point)
a) Dwight Howard
b) Brook Lopez
c) Kevin Durant
d) Russell Westbrook

Week 3 — Quiz 2

Question 1

In the two games between Kings XI Punjab and Delhi Daredevils, which team batted first in each game? (1 point)
a) Delhi in both games
b) Delhi in the first game, Kings XI in the second game
c) Kings XI in both games
d) Kings XI in the first game, Delhi in the second game

Question 2

In the two games between Kings XI Punjab and Delhi Daredevils, (1 point)
a) Delhi won the first game and Kings XI won the second game
b) Kings XI won the first game and Delhi won the second game
c) Kings XI won both games
d) Delhi won both games

Question 3

In the first game, (1 point)
a) Kings XI had the faster run rate and Delhi XI lost fewer wickets
b) Delhi had the faster run rate and lost fewer wickets
c) Kings XI had the faster run rate and lost fewer wickets
d) Delhi had the faster run rate and Kings XI lost fewer wickets

Question 4

By the midpoint of each inning (60 balls) In the second game between Kings XI Punjab and Delhi Daredevils, (1 point)
a) Delhi had fewer runs and Kings XI had lost more wickets
b) Delhi had fewer runs and had lost more wickets
c) Kings XI had fewer runs and Delhi XI had lost more wickets
d) Kings XI had fewer runs and had lost more wickets

Question 5

Comparing the two games, which game had the higher numbers of runs per wicket? (1 point)
a) The first game (game 2)
b) The second game (game 22)
c) About the same
d) Impossible to tell from the plots

Week 4 — Quiz 1

Question 1

Which of the following about the NFL_Game dataframe is incorrect? (1 point)
a) The dataframe has 24,314 observations and 28 variables.
b) The highest score achieved by an away team was 72.
c) The dataframe covers games from 1966 to 2019 (1966-1967 season to 2019-2020 season).
d) NFL teams earned a higher average score at home than away.

Question 2

Which of the following statements regarding the correlation coefficients calculated in step 3 is *incorrect*? (1 point)
a) The correlation coefficient between “score” and “weather_wind_mph” is -0.079. This means there is a weak negative relationship between a team’s final score and wind speed.
b) The correlation coefficient between “win” and “home” is 0.15. This means that teams won more at home games than away games.
c) The correlation coefficient between “score” and “weather_temperature” is -0.03. This means that teams earned slightly lower scores when the temperature got higher.
d) The correlation coefficient between “score_diff” and “home” is 0.17. This means that the difference between own score and opponent’s score tend to be higher for home team than away team.

Week 4 — Quiz 2

Question 1

Which of the following is a correct interpretation of the regression results of Reg1_1? (1 point)

a) The estimated coefficient of the “home” variable is 5.22. This means that the score difference between a team and its opponent is about 5.22 higher for a home team than an away team.

b) The R-squared of this regression is 0.03. This means that there is no home game advantage for the NFL teams.

c) The t-statistics of the estimate on the “home” variable is 27.307. This suggests that this estimate is not statistically significant.

d) The R-squared of this regression is 0.03. This means that there is a strong linear relationship between “home” and “score_diff.”

Question 2

What are the estimated coefficient and the statistical significance for the “stadium_neutral” variable from Reg1_2? (1 point)

a) 7.25; not statistically significant

b) 7.25; statistically significant at 0.01 level.

c) -14.50; statistically significant at 0.01 level

d) -14.50; not statistically significant

Question 3

Which of the following is a correct interpretation of the **impact of being a home team *and* the impact of playing at a neutral stadium** based on the results of Reg1_3? (1 point)

a) For home teams, playing at a neutral stadium decreases its score difference by 6.74.

b) For home teams, playing at a neutral stadium decreases the score difference by 8.16.

c) Playing at a neutral increases the difference between own score and opponent’s score by 6.74.

d) Playing at home increases the difference between own score and opponent’s score by 5.33.

Week 4 — Quiz 3

Question 1

Which of the following statement regarding the result of Reg2_1 is *incorrect*? (1 point)
a) The estimated coefficient on the “home” variable is 2.61. This means that in each game, home team is estimated to score 2.61 more than the away team.
b) The estimate of the “season” variable is 0.05 and it is statistically significant at 0.01 level.
c) The estimated coefficient of home game is 2.61 and it is statistically significant at 0.01 level.
d) The estimated coefficient on the “season” variable is 0.05. This means that each season, the average score earned by each team in each game increases by 0.05 compared to the previous season.

Question 2

Based on the results from Reg2_2, which of the following factor *does not* affect scores in a *linear* fashion? (1 point)
a) Temperature
b) Humidity
c) Season
d) Wind

Question 3

Which of the following interpretations of the results from regression Reg2_3 is *incorrect?* (1 point)
a) The estimated coefficient of stadium capacity is
-1.24×10^{-5}.
Since the size of the estimate is very small, it is not statistically significant.

b) Humidity has a negative impact on team’s final score. When the relative humidity increases by 1%, a team will score 0.014 less.
c) Wind has a negative impact on team’s score. When wind speed increases by 1 mph, a team will earn 0.13 less in final score.
d) The estimated coefficient on the age of the stadium is 0.0053 but it is not statistically significant at 0.1 level.

Question 4

Based on the results from regression Reg2_4, which of the following correctly describes the effect of stadium type on individual team's final score? (1 point)

a) The estimated effect of retractable stadium is 0.92 which suggests that teams score 0.92 more points at retractable stadiums than at indoor stadiums.

b) The estimated effect of retractable stadium is 0.92 which suggests that teams score 0.92 more at retractable stadiums than other stadiums.

c) The estimated effect of outdoor stadium is 0.50 which suggests that outdoor stadium is associated with higher final scores.

d) Both the estimated coefficients on the "stadium_type" variables are positive which means that there is a positive relationship between stadium type and score.

Week 5 — Quiz 1

Question 1

What was the value of the sum of all salaries in 2014? (1 point)

A. $82 million
B. $65 million
C. $84 million
D. $74 million

Question 2

In the first regression, what is the coefficient on relsal? (1 point)

A. 0.2050
B. 0.4752
C. 0.403
D. 0.509

Question 3

Based on the first regression, which of the following statements is the most accurate? (1 point)

A. The coefficient of relsal is insignificant
B. The regression implies that increasing salaries causes higher win percentage
C. The coefficient of relsal is insignificantly different from zero at the 5% level (p-value)
D. We cannot say if relsal significantly affects win percentage

Question 4

In the second regression, which of the following statements is **true**: (1 point)

A. Neither relsal nor lagged win percentage are statistically significant at the 5% level
B. Both relsal and lagged win percentage are statistically significant at the 5% level
C. Lagged win percentage is statistically significant at the 5% level but relsal is not
D. Relsal is statistically significant at the 5% level but lagged win percentage is not

Question 5

Based on the second regression, which of the following statements is the most accurate? (1 point)

A. Relsal cannot explain win percentage
B. The regression suggests that relsal is more important than lagged wpc in explaining performance
C. The addition of lagged win percentage has not significantly improved on the explanatory power of the regression compared to the first regression
D. The regression is an improvement on the first regression

Question 6

The third regression includes fixed effects. Not every team in a regression can have a fixed effect, there must be one “reference” team, relative to which each fixed effect is defined. Thus each fixed effect listed should be understood as meaning the performance of the team relative to the reference team. In this case, what is the reference team? (1 point)

A. Chennai Super Kings
B. Kochi Tuskers Kerala
C. Sunrisers Hyderabad
D. Deccan Chargers

Question 7

Ignoring Rising Pune Supergiants (for which there is only one observation) all of the fixed effects are negative. What does this tell us about the reference team?? (1 point)

A. All else equal, it is the worst team in the league.
B. The team has the lowest salaries in the league
C. The team has the highest salaries in the league
D. All else equal, it is the best team in the league.

Question 8

Looking at the coefficient and standard error for the Mumbai Indians fixed effect, which would you say best describes the statistical inference that can be drawn? (1 point)

A. Mumbai is statistically no better or worse than the reference team
B. Mumbai is statistically better than the reference team
C. Mumbai is statistically the second best team in the IPL
D. Mumbai is statistically worse than the reference team

Question 9

Based on the third regression, which of the following statements is the most accurate? (1 point)

A. The addition of fixed effect has not significantly improved on the explanatory power of relsal or lagged win percentage compared to the first two regressions
B. The negative coefficient on relsal suggests that higher salary spending leads to lower win percentage
C. The negative coefficient on lagged win percentage suggests that higher win percentage last year leads to lower win percentage this year
D. The increased R2 means that we should have more confidence in the coefficient value in this regression than the previous two regressions

Question 10

Based on these plots, which of the six players seems most likely to found shooting from the left hand side of the basket rather than the right hand side? (1 point)

A. The market for cricketers is not efficient
B. Salaries are not measured accurately and so the true value of relsal is not being tested
C. Because there is an effective salary cap, salaries do not in fact vary enough to make a difference
D. Players play for the love of the game and not for money

Week 6 — Quiz 1

Question 1

Which of the following statements about the 2014-2015 season NBA data is *incorrect*? (1 point)

A. There is information on 120 games in the shotlog dataframe.
B. There is missing value in the “shot_clock” variable.
C. There is information on 281 NBA players.
D. The shotlog dataframe covers information on 128,069 shots.

Question 2

Which of the following statements regarding the prediction errors calculated in step 4 is *incorrect*? (1 point)

A. The maximum for both prediction errors is 0.69
B. The average current period prediction error is -1.099
C. There are 113,726 observations with real value in the prediction error for the previous period.
D. The standard deviation for the current period prediction error is 0.49.

Week 6 — Quiz 2

Question 1

What is the mean of the conditional probability for players to make a shot given that they made the previous one? (1 point)

A. 0.45
B. 0.38
C. 0.06
D. 0.18

Question 2

Which of the following statements is a correct interpretation of the t-test performed in step 6? (1 point)

A. The p-value for the test is 6.93 x 10^-38. We don't have enough evidence to suggest the conditional probability is statistically significantly different than the unconditional probability.
B. The p-value for the test is 6.93 x 10^-38. This suggests that the conditional probability is statistically significantly different than the unconditional probability.
C. The test statistics is -13.89. We don’t have enough evidence that the conditional probability is statistically significantly different than the unconditional probability.
D. The test statistics is -13.89. This means that the conditional probability is statistically significantly less than the unconditional probability.

Question 3

What is the first order autocorrelation coefficient for successfully making a shot calculated in step 7? (1 point)

A. -0.01
B. 0.75
C. 0.97
D. 0.92

Question 4

Which of the following four players has the highest first order autocorrelation coefficient for making a shot? (1 point)

A. Jason Smith
B. Cole Aldrich
C. Mike Miller
D. Alex Len

Week 6 — Quiz 3

Question 1

What is the first order autocorrelation coefficient for successfully making a shot calculated in step 7? (1 point)

A. The estimated coefficient on the “lagerror” variable is statistically significant at 0.1 level with a p-value of 0.071 in Reg1. Therefore, we have found evidence of “hot hand.”
B. The estimated coefficient of the “closet_def_dist” variable is positive in both regressions. This means that the closer the closest defender is, the more likely a player performs better than expected in the given shot.
C. The estimated coefficient of the “points[T.3]” variable is positive in both regressions. This means that compared to 3-point field goals, players are more likely to perform better than expected in 2-point field goals.
D. Neither regression provides evidence of “hot hand.”

Question 2

Based on the results from Reg3_player, which of the following players’ shooting success rate is affected by the distance of the shot the most? (1 point)

A. James Harden
B. Andrew Wiggins
C. Russel Westbrook
D. Stephen Curry

Question 3

What is the first order autocorrelation coefficient for successfully making a shot calculated in step 7? (1 point)

A. Stephen Curry
B. Alonzo Gee
C. Cole Aldrich
D. Reggie Jackson

Jottings

With **Autocorrelation,** in particular the first autocorrelation, we test the linear relationship between the outcome of adjacent time periods, such as the relationship between performance this season and performance last season. The relationship between, scoring the current goal and scoring the previous goal.

Weighted least squares is an estimation technique that weights the observations proportional to the reciprocal of the error variance for the observation. Therefore overcomes the issue of non-constant variance.