



PROJECT-SET

Statistics Education for Teachers

NFL Quarterback Salaries

Overview of Lesson

In this activity, students will set up a statistical question to explore how to interpret a linear regression equation and the correlation coefficient for a relationship between two quantitative variables. Students are divided into groups of three with each student being presented with a table of data showing statistics about the 30 top paid NFL quarterbacks in 2009-10. The data include total salary, pass completion percentage, total number of touchdowns, and average number of yards per game for each quarterback.

Each person in the group will receive data about the 30 highest paid NFL quarterback salaries in 2009-2010 and the other three variables listed above. Using these data, students set up a statistical question concerning the salary of a top paid NFL quarterback in 2009-2010 and the explanatory variable on their sheet. They use technology to create a scatterplot with the graph of the regression line superimposed and find a linear regression equation along with the correlation r and the coefficient of determination r^2 . The three members of each group then interpret and compare their results and decide which of the three explanatory variables seemed to be the best predictor of a top paid 2009-2010 NFL quarterback's salary. The lesson includes questions that require students to demonstrate an understanding of the concept of correlation versus causation.

GAISE Components

This activity follows four components of statistical problem solving put forth in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report*. The four components are: Formulate Question, Collect Data, Analyze Data, and Interpret Results. This is a GAISE level C lesson which incorporates the following process components:

- I. Formulate Question
 - Students pose their own questions of interest
- II. Collect Data
 - Students make *design for differences*
- III. Analyze Data
 - Quantification of association; fitting of models for association
- IV. Interpret Results
 - Interpret measures of strength of association
 - Interpret models of association

Common Core State Standards for Mathematical Practice

4. Model with mathematics.
7. Look for and make use of structure.

Common Core State Standards Grade Level Content

• High School

S-ID **Interpreting Categorical and Quantitative Data**

Summarize, represent, and interpret data on two categorical and quantitative variables

6. Represent data on two quantitative variables on a scatter plot and describe how the variables are related.
 - a. Fit a function to the data; use functions fitted to data to solve problems in the context of the data. *Use given functions or choose a function suggested by the context. Emphasize linear, quadratic, and exponential models.*
 - b. Fit a linear function for a scatter plot that suggests a linear association. *Interpret linear models.*
7. Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.
8. Compute (using technology) and interpret the correlation coefficient of a linear fit.

NCTM Principles and Standards for School Mathematics

Data Analysis and Probability Standards for Grades 9-12

Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them

- understand the meaning of measurement data and categorical data, of univariate and bivariate data, and of the term variable
- understand histograms, parallel box plots, and scatterplots and use them to display data

Select and use appropriate statistical methods to analyze data

- for bivariate measurement data, be able to display a scatterplot, describe its shape, and determine regression coefficients, regression equations, and correlation coefficients using technological tools
- identify trends in bivariate data and find functions that model the data or transform the data so that they can be modeled

Prerequisites

Prior to completing this activity, students should be able to set up a statistical question that explores the relationship between two quantitative variables. They should have practiced defining the population and relationship of interest and have experience with both determining and carrying out a data sampling method. They should be able to determine whether the data allows one to estimate causal effects. Students should have previous experience with the following bivariate data analysis techniques:

- plotting sample data in a scatterplot and visually determining and describing the trend of the scatterplot
- examining scatterplots and placing a model on the graph (e.g., drawing a line or curve)
- discussing where the “line of best fit” would be placed when a linear model is appropriate
- discuss mathematical ways with which the relationship could be represented/modeled
- finding the sample regression equation for a sample using technology

Students should have practiced the following as they interpreted the results of bivariate data analysis:

- interpreting the slope and y-intercept in context of the posed question for the regression line

- showing the residuals on the scatterplot by noting the vertical distance between the predicted value on the regression line and the actual data point
- distinguishing between positive and negative residuals
- obtaining residuals through technology and having technology show the residual plot
- measuring the variability in the residuals using $\sum(y - \hat{y})^2$
- informally describing correlation by examining how the points cluster
- showing that the point (\bar{x}, \bar{y}) is always on the regression line
- interpreting the residual plot, relating it back to the residuals seen on the scatterplot
- showing the residuals sum to 0
- understanding how ordinary least squares (OLS) can be viewed as minimizing the sum of the residuals

Learning Targets

In this lesson, students will gain continued experience with setting up a statistical question that explores a relationship between two quantitative variables. They will determine and carry out a data sampling method, using bivariate data analysis techniques and interpretation of results. In addition to the prerequisite skills listed above, students will gain an understanding of what correlation represents on a scatterplot. In their interpretation of the data, they will learn what r and r^2 represent and be able to explain, in context of the posed question, that r^2 is the fraction of the sample variance explained by the explanatory variable x .

Time Required

The time required for this activity is roughly 90 minutes.

Materials Required

For this activity, students will need a pencil, paper, “What Makes NFL Quarterbacks Worth Their Salaries?” activity sheet, and a graphing calculator. Note: the powerpoint for this lesson is available on the ProjectSET website.

Instructional Lesson Plan

The GAISE Statistical Problem-Solving Procedure

1. Formulate Question(s)

Before beginning the activity, the class will be divided into groups of three. Each student in the class will then be given an introductory activity sheet with a table of data showing the salary, pass completion percentage, total number of touchdowns, and average number of yards per game for the 30 top paid NFL quarterbacks in 2009-2010. Two introductory questions on the sheet will prompt students to observe the wide range of quarterback salaries and consider what factors might determine how much an NFL quarterback is paid. These questions can be discussed as a class or within each group.

After the discussion, the teacher will distribute the activity sheets to each group. Each activity sheet will contain data and questions about the relationship between quarterback salaries and the three possible explanatory variables. The first question on the student activity sheet instructs students to review the data and write a statistical question concerning the salary of a top paid NFL quarterback in 2009-2010 and the explanatory variable on their sheet. Possible student questions may include: 1) Is there an association between the 30 top NFL quarterback salaries in 2009-2010 and the quarterback's pass completion percentage, 2) Can we use the total number of touchdowns of the 30 top paid NFL quarterbacks in 2009-2010 to predict their salaries, or 3) Is there an association between the 30 top NFL quarterback salaries in 2009-2010 and the quarterback's average number of yards per game?

II. Collect Data

After formulating the question to be answered, students will be given specific directions for entering the bivariate data on their activity sheet in their graphing calculator. The data for the explanatory and response variables of interest are thus previously collected for the students. There are two activity sheets available for this lesson. If students use graphing calculators, they will be given specific instructions (see NFL Activity Sheet below) about how to enter the data in Lists 1 and 2. If students use other software, they can use the non-calculator activity sheet to carry out this lesson plan. Regardless of how the activity is completed, throughout the process they will be prompted to distinguish between the explanatory and response variables.

III. Analyze the Data

After entering the data in Lists 1 and 2 on their graphing calculators or other software, students will be required to define and graph a scatterplot in the statistics plot menu and sketch it on the activity sheet. This process will require them to consider how to best represent the data on a scatterplot, including making decisions about how to scale and label axes for the explanatory and response variables.

Then, students will be given detailed instructions for using the data in Lists 1 and 2 (if this is your chosen method) to determine the least-squares regression equation, correlation r , and coefficient of determination, r^2 . They will transfer this information from their calculator to the activity sheet and use it to perform further analysis of the linear relationship between the explanatory and response variables. Included in the directions for finding the least-squares regression equation will be instructions for entering the equation in the graph menu so the least-squares line can be added to the scatterplot. Students will be instructed to add this line to the sketch of the scatterplot on their activity sheet.

CORRECT RESPONSES FOR DATA ANALYSIS

Pass Completion vs. Salary:

Sample Regression Equation: $\hat{y} = -4,443,755.852 + 244,351.3169x$

\hat{y} = predicted salary of 2009-2010 top paid NFL quarterback in \$

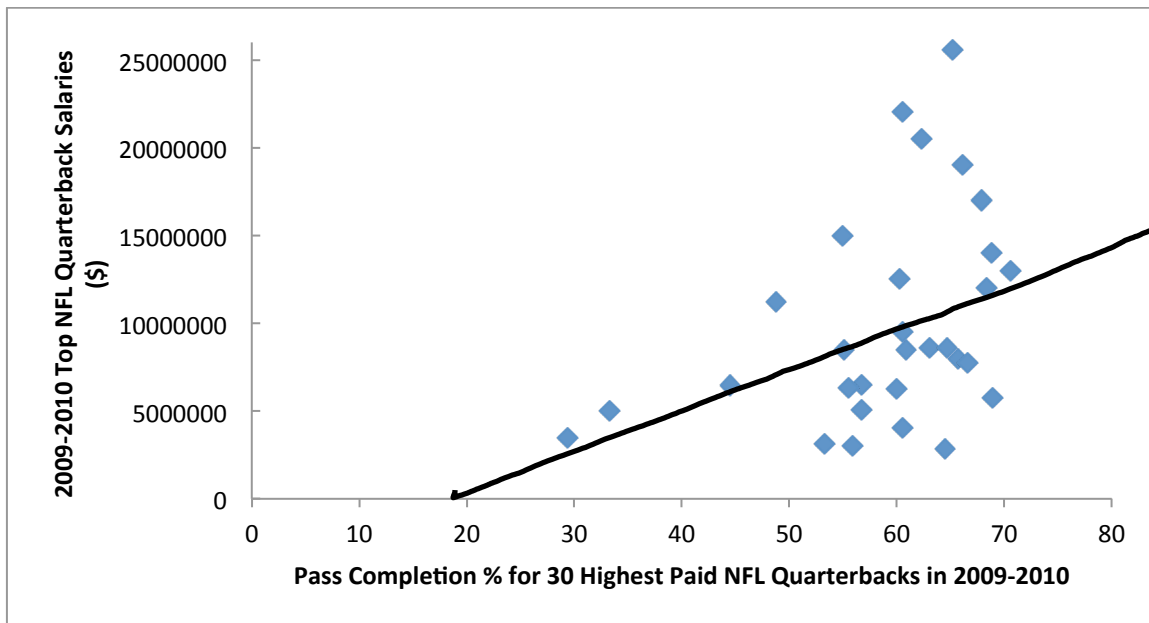
x = pass completion % for top paid NFL quarterback in 2009-2010

It is important that students use \hat{y} (this is because it represents the predicted value from the regression equation, not the actual value of the data) instead of y for the response variable and define it to be the **predicted** salary of 2009-2010 top paid NFL quarterback in dollars (\$).

Correlation r : $r = 0.3940$

Coefficient of Determination r^2 : $r^2 = 0.1553$

Scatterplot with Least-Squares Regression Line:



It is important that students label the two axes, clearly indicating the explanatory variable on the horizontal axis and the response variable on the vertical axis.

Number of Touchdowns vs. Salary

Sample Regression Equation: $\hat{y} = 4,393,649.841 + 320,510.258x$

\hat{y} = predicted salary of 2009-2010 top paid NFL quarterback in \$

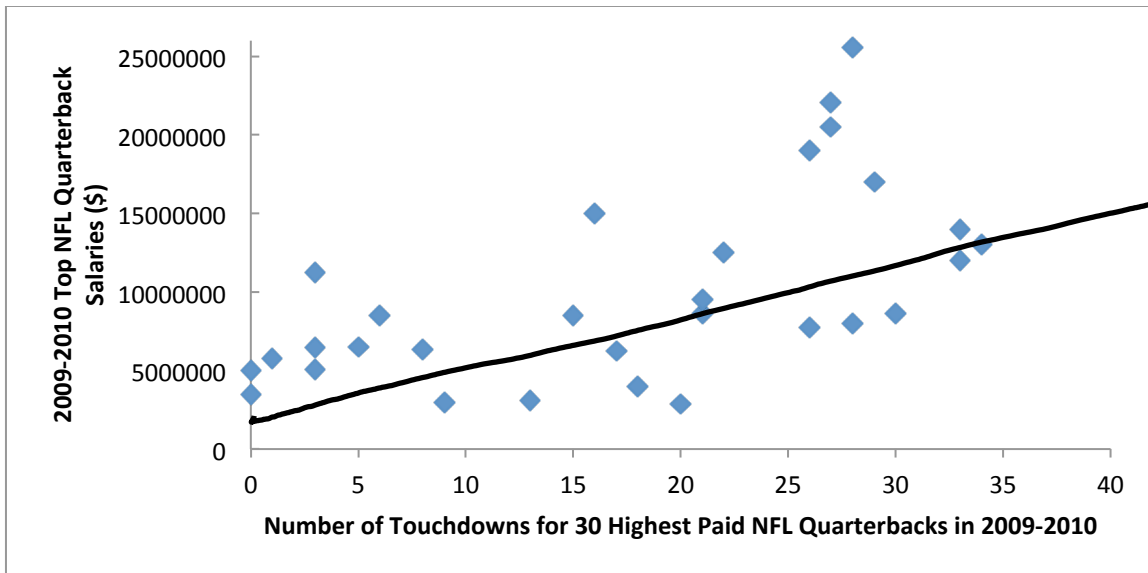
x = total number of touchdowns in 2009-10 of top paid NFL quarterback

It is important that students use \hat{y} instead of y for the response variable and define it to be the **predicted** salary of 2009-2010 top paid NFL quarterback in \$.

Correlation r : $r = 0.5954$

Coefficient of Determination r^2 : $r^2 = 0.3545$

Scatterplot with Least-Squares Regression Line: (on next page)



It is important that students label the two axes, clearly indicating the explanatory variable on the horizontal axis and the response variable on the vertical axis.

Average Number of Yards per Game vs. Salary

Sample Regression Equation: $\hat{y} = 2,242,948.153 + 38,047.530x$

\hat{y} = predicted salary of 2009-2010 top paid NFL quarterback in \$

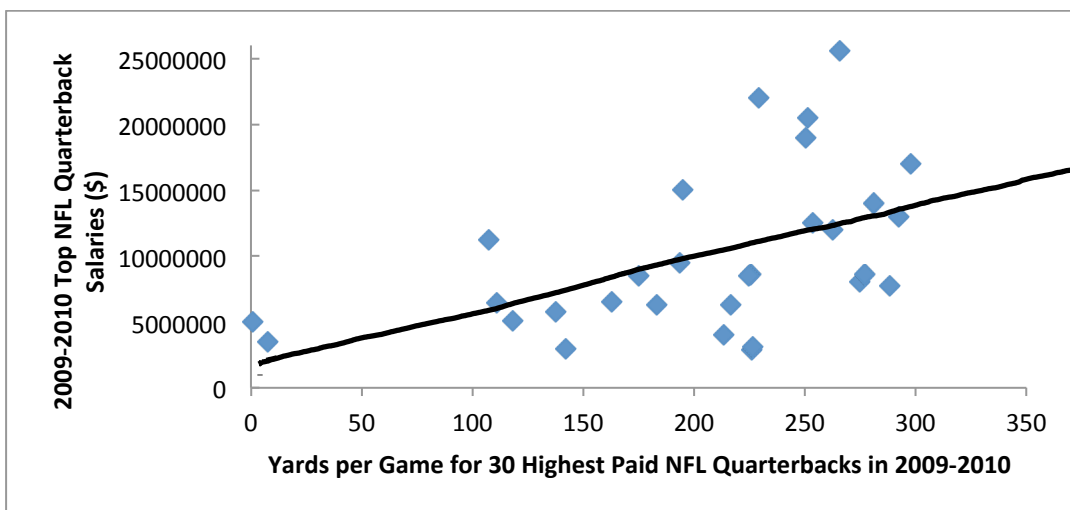
x = yards per game in 2009-2010 for top paid NFL quarterback

It is important that students use \hat{y} instead of y for the response variable and define it to be the **predicted** salary of 2009-2010 top paid NFL quarterback in \$.

Correlation r : $r = 0.4886$

Coefficient of Determination r^2 : $r^2 = 0.2387$

Scatterplot with Least-Squares Regression Line:



It is important that students label the two axes, clearly indicating the explanatory variable on the horizontal axis and the response variable on the vertical axis.

IV. Interpret the Results

During the activity, students were asked to write a statistical question concerning the salary of a top paid NFL quarterback in 2009-2010 and a given explanatory variable (pass completion percentage, total number of touchdowns, or yards per game). They perform an analysis of data they are given, which includes finding the correlation, r , and coefficient of determination, r^2 . Each student is asked to give an interpretation of the correlation, r , and coefficient of determination, r^2 they have found. The correct student responses are given below.

Pass Completion vs. Salary

Correlation r : $r = 0.3940$

Interpretation: There is a weak to moderate positive linear relationship between the pass completion percentage of the 30 top paid NFL quarterbacks in 2009-2010 and their salaries. Students need to include the strength (weak to moderate), direction (positive), and form (linear) in order for the answer to be complete.

Coefficient of Determination r^2 : $r^2 = 0.1553$

Interpretation: 15.53% of the variation in the salaries of the 30 top paid NFL quarterbacks in 2009-2010 is explained by the straight-line relationship between their pass completion percentage and salaries. This means that 84.47% of the variation in salaries is explained by factors other than the quarterbacks' pass completion percentages.

Number of Touchdowns vs. Salary

Correlation r : $r = 0.5954$

Interpretation: There is a moderate positive linear relationship between the total number of touchdowns of the 30 top paid NFL quarterbacks in 2009-2010 and their salaries. Students need to include the strength (moderate), direction (positive), and form (linear) in order for the answer to be complete.

Coefficient of Determination r^2 : $r^2 = 0.3545$

Interpretation: 35.45% of the variation in the salaries of the 30 top paid NFL quarterbacks in 2009-2010 is explained by the straight-line relationship between their total number of touchdowns and their salaries. This means that 64.55% of the variation in salaries is explained by factors other than their total number of touchdowns.

Average Number of Yards per Game vs. Salary

Correlation r : $r = 0.4886$

Interpretation: There is a moderate positive linear relationship between the total number of touchdowns of the 30 top paid NFL quarterbacks in 2009-2010 and their salaries. Students need to include the strength (moderate), direction (positive), and form (linear) in order for the answer to be complete.

Coefficient of Determination r^2 : $r^2 = 0.2387$

Interpretation: 23.87% of the variation in the salaries of the 30 top paid NFL quarterbacks in 2009-2010 is explained by the straight-line relationship between their yards per game and their salaries. This means that 76.13% of the variation in salaries is explained by factors other than their yards per game.

After interpreting the results of their analysis, students are asked to consider a hypothetical situation in which the correlation had instead been $r = 0.98$. Then they are asked to answer the following two questions:

- (1) What would this value of r tell you about the nature of the association between the salaries of the top paid NFL quarterback in 2009-2010 and (*insert: the specific explanatory variable considered by the student*)?

Sample Answer: A value of $r = 0.98$ would tell us that there is a strong, positive, linear association between the salaries of the 30 top paid NFL quarterbacks in 2009-2010 and (*insert: the specific explanatory variable considered by the student*).

- (2) Would this correlation value of $r = 0.98$ have been evidence that a quarterback's high pass completion percentage caused his salary to increase? Why or why not?

Sample Answer: The strong association would not have been evidence that a quarterback's large number of touchdowns caused his salary to increase. Although a correlation of $r = 0.98$ would indicate a strong linear relationship between a quarterback's number of touchdowns and his salary, we cannot conclude that an increase in a quarterback's salary is **caused** by his large number of touchdowns. There could be other variables that contribute to the relationship between the two variables. A strong association between two variables is not enough to draw conclusions about cause and effect. Student answers will vary, but need to clearly indicate that association does not imply causation.

The final question on the activity sheet asks the three members of each group to compare their results and decide which of the three explanatory variables seems to be the best predictor of a 2009-2010 top paid NFL quarterback's salary. Students should conclude that the quarterback's total number of touchdowns seems to be the best predictor of a top paid 2009-2010 NFL quarterback's salary and should give the following justifications for their answer.

- (1) The correlation $r = 0.5954$ for the linear relationship between the salary of a top paid NFL quarterback in 2009-2010 and the quarterback's total number of touchdowns is higher than the correlation r -values for the linear relationships between the salary of a top paid NFL quarterback in 2009-2010 and the other two explanatory variables. This tells us that the linear relationship between the salary and total number of touchdowns is stronger than the linear relationships between the salary and the other two explanatory variables.

- (2) The coefficient of determination for the linear relationship between the salary of a top paid NFL quarterback in 2009-2010 and the quarterback's total number of touchdowns is $r^2 = 0.3545$. This is higher than the r^2 values for quarterback salaries versus the other two explanatory variables.

This tells us that a larger percentage of the variation in the salaries of the 30 top paid NFL quarterbacks in 2009-2010 is explained by the straight-line relationship between their total number of touchdowns and their salaries than by the straight-line relationships between the other two explanatory variables and their salaries.

Assessment

1. The correlation r measures what two characteristics of the linear association between two quantitative variables?

2. How does one use the correlation r to determine the direction of the linear association between two quantitative variables?
3. How does one use the correlation r to determine the strength of the relationship between two quantitative variables?
4. Explain what the coefficient of determination r^2 tells you about how well a regression line fits a set of data.

Answers to Assessment

1. The correlation r measures the direction and strength of the linear association between two quantitative variables.
2. $r > 0$ indicates a positive association and $r < 0$ indicates a negative association.
3. The correlation r always takes on values $-1 \leq r \leq 1$ and indicates the strength of a relationship by how close it is to -1 or 1 .
4. The coefficient of determination, r^2 , tells us the fraction of the variation in the values of the response variable y that is accounted for by the least-squares regression line of y on the explanatory variable x .

Possible Extensions

1. After students have mastered the skills necessary to interpret a linear regression equation and the correlation for a relationship between two quantitative variables measured on an entire population, they can explore the relationship between two quantitative variables for which you **cannot** get data for the entire population. Then, they would use the concept of random sampling from the population of interest to obtain a sample regression equation and correlation as estimates of population parameters. This would then be extended to estimating the true regression equation from a population from which multiple random samples are taken using the regression equation computed from the repeated random samples.

References

1. *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report*, ASA, Franklin et al., ASA, 2007 <http://www.amstat.org/education/gaise/>.
2. www.usatoday.com/sportsdata/football/nfl/salaries/position/QB/2009
3. www.pro-football-reference.com/years/2009/passing.htm

Activity Sheet

What Makes NFL Quarterbacks Worth Their Salaries?

Introduction

How does an NFL team decide how much to pay the quarterback? What makes one quarterback worth a total salary of \$25,556,630 and another only around \$3,000,000? The data below show statistics about the 30 top paid NFL quarterbacks in 2009-10. Their total salary is shown, along with their pass completion percentage, total number of touchdowns, and average number of yards per game.

2009-10 Statistics for 30 Highest Paid NFL Quarterbacks

Player	Salary	Pass Completion	Touchdowns	Yards per
		%		Game
1. Derek Anderson	6,450,000	44.5	3	111
2. Tom Brady	8,007,280	65.7	28	274.9
3. Drew Brees	12,989,500	70.6	34	292.5
4. Marc Bulger	6,507,280	56.7	5	163.2
5. Jason Campbell	2,864,780	64.5	20	226.9
6. Matt Cassel	15,005,200	55	16	194.9
7. Kerry Collins	8,507,280	55.1	6	175
8. Daunte Culpepper	5,050,000	56.7	3	118.1
9. Jay Cutler	22,044,090	60.5	27	229.1
10. Jake Delhomme	6,325,000	55.5	8	183.2
11. Brett Favre	12,000,000	68.4	53	262.6
12. Ryan Fitzpatrick	2,995,590	55.9	9	142.2
13. Joe Flacco	8,601,760	63.1	21	225.8
14. David Garrard	8,500,000	60.9	15	224.8
15. Matt Hasselbeck	6,256,240	60	17	216.4
16. Eli Manning	20,500,000	62.3	27	251.3
17. Peyton Manning	14,005,720	68.8	33	281.3
18. Luke McCown	5,006,760	33.3	0	0.7
19. Donovan McNabb	12,507,280	60.3	22	253.8
20. Carson Palmer	9,500,000	60.5	21	193.4
21. Chad Pennington	5,750,000	68.9	1	137.7
22. Philip Rivers	25,556,630	65.2	28	265.9
23. Aaron Rodgers	8,600,000	64.7	30	277.1
24. Ben Roethlisberger	7,751,560	66.6	26	288.5
25. JaMarcus Russell	11,255,440	48.8	3	107.3
26. Matt Schaub	17,000,000	67.9	29	298.1
27. Chris Simms	3,466,500	29.4	0	7.7
28. Alex D. Smith	4,007,280	60.5	18	213.6
29. Matthew Stafford	3,100,000	53.3	13	226.7
30. Kurt Warner	19,004,680	66.1	27	250.2

Data was taken from the following sources:

www.pro-football-reference.com/years/2009/passing.htm

www.usatoday.com/sportsdata/football/nfl/salaries/position/QB/2009

Problem

Use the data collected to investigate the relationship between the 30 top NFL quarterback salaries in 2009-2010 and the following three variables:

1. Pass Completion Percentage in 2009-2010
2. Total Number of Passing Touchdowns in 2009-2010
3. Average Number of Yards per Game in 2009-2010

Instructions

Your class will be divided into groups of three. Each person in the group will receive data about the 30 highest paid NFL quarterback salaries in 2009-2010 and one of three other variables listed above. Using the data, your group will explore the relationship between the salary of a 2009-2010 top paid NFL quarterback and each of these variables.

1. Determine a Statistical Question That Involves the Data Given

After reviewing the data, write a statistical question concerning the salary of a top paid NFL quarterback in 2009-2010 and your variable.

What is the population of interest in your question?

What is the relationship of interest?

2. Make a Scatterplot to View the Relationship Between the Two Variables

Sketch the graph below. Make sure that you scale and label the axes



3. Find Sample Regression Equation Using Technology

Write the least-squares regression equation for the sample you found defining the variables in the context of the problem.

4. Graph the Regression Line on your Scatterplot

5. Find and Interpret Correlation r

What does the value of r tell you about the nature of the association between the salary of a top paid NFL quarterback in 2009-2010 and your variable?

6. Find and Interpret the Coefficient of Determination r^2

Write the value for r^2 below and explain what this tells you about the association between the salary of a top paid NFL quarterback in 2009-2010 and your variable.

7. Correlation vs. Causation

(a) In #6, you found a value for the correlation r that indicated a weak to moderate positive, linear relationship between the pass completion % of the 30 top paid NFL quarterbacks in 2009-2010 and their salaries.

Suppose the correlation had been $r = 0.98$. What would this value of r tell you about the nature of the association between the salaries of top paid NFL quarterback in 2009-2010 and your variable?

(b) As you have just seen, if the correlation had been $r = 0.98$, there would have been a very strong association between the salaries of top paid NFL quarterbacks in 2009-2010 and their pass completion percentages. Would this have been evidence your variable **caused** his salary to increase? Why or why not?

Activity Sheet with Calculator Instructions (example for Number of Touchdowns vs. Salary)

1. Determine a Statistical Question That Involves the Data Given

After reviewing the data, write a statistical question concerning the salary of a top paid NFL quarterback in 2009-2010 and the quarterback's total number of touchdowns.

What is the population of interest in your question?

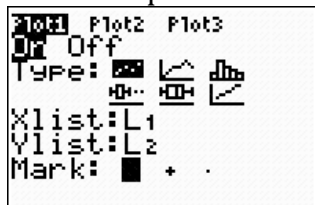
What is the relationship of interest?

Enter Data Values Into Lists on Graphing Calculator

Clear lists L1 (List 1) and L2 (List 2) on your graphing calculator. Enter the number of touchdowns in L1 and the quarterback salaries in L2. The number of touchdowns in L1 will be the explanatory variable and the quarterback salaries in L2 will be the response variable.

Using Data Entered on #2, Make a Scatterplot on Graphing Calculator

- Define a scatterplot in the statistics plot menu. Specify the settings shown.



- Use ZoomStat to obtain a graph. The calculator will set the window dimensions automatically by looking at the values in L1 and L2.

2. Make a scatterplot to view the relationship between the two variables.

Sketch the graph below. Make sure that you scale and label the axes. (Hint: You can use TRACE on your calculator to help you label the axes.)



3. Find Sample Regression Equation Using Graphing Calculator

To determine the least-squares regression equation for the data in L1 and L2, carry out the following commands on your graphing calculator:

- Press the STAT key; choose CALC and then 8:LinReg(a + bx).
- Finish the command to read LinReg(a + bx) L1, L2, Y1 and press ENTER.

Write the least-squares regression equation you found on your calculator:

4. Graph the Regression Line on Graphing Calculator

Turn off all other equations in the Y= screen and press GRAPH to add the least-squares line to the scatterplot. Add this line to the sketch of your scatterplot on #3.

5. Find and Interpret Correlation r

To determine the correlation r for the linear relationship, you will again use the data entered in L1 and L2. You will carry out the following commands as you did on #4.

- Press the STAT key; choose CALC and then 8:LinReg(a + bx).
- Finish the command to read LinReg(a + bx) L1, L2, Y1 and press ENTER.

Write the value of the correlation r that you will see at the bottom of the list on the screen. What does the value of r tell you about the nature of the association between the salary of a top paid NFL quarterback in 2009-2010 and the quarterback's total number of touchdowns?

NOTE: If the value of r does not appear on the screen, enter 2ND 0 to access the catalog on your calculator. Scroll down to DiagnosticOn and press ENTER. Press ENTER a second time. You should see the word "Done" on your homescreen, verifying that you have turned on the diagnostics. Now the value of the correlation r should appear on the screen every time you calculate the linear regression equation.

6. Find and Interpret the Coefficient of Determination r^2

In #6, you also saw a value for r^2 on the screen on your calculator with the linear regression equation. (You could also find this value by squaring the value for r you found in #6.) Write the value for r^2 below and explain what this tells you about the association between the salary of a top paid quarterback in 2009-2010 and the quarterback's total number of touchdowns.

7. Correlation vs. Causation

- (a) In #6, you found a value for the correlation r that indicated a moderate positive, linear relationship between the number of touchdowns of the 30 top paid NFL quarterbacks in 2009-2010 and their salaries.

Suppose the correlation had instead been $r = 0.98$. What would this value of r tell you about the nature of the association between the salary of a top paid NFL quarterback in 2009-2010 and the quarterback's number of touchdowns?

- (b) As you have just seen, if the correlation had been $r = 0.98$, there would have been a very strong association between the salary of a top paid NFL quarterback in 2009-2010 and the quarterback's number of touchdowns. Would this have been evidence that a quarterback's large number of touchdowns caused his salary to increase? Why or why not?

8. After the three members of your group have completed #1-8, compare your results and answer the following question.
- Based on the results of your investigations, which of the three explanatory variables seemed to be the best predictor of a top paid 2009-2010 NFL quarterback's salary? **Justify your answer.**

Answers to Activity Sheet for Pass Completion vs. Salary

1. **Statistical Question** - Answers will vary, but should be similar to the following:

Is there an association between the 30 top NFL quarterback salaries in 2009-2010 and the quarterbacks' pass completion percentages?

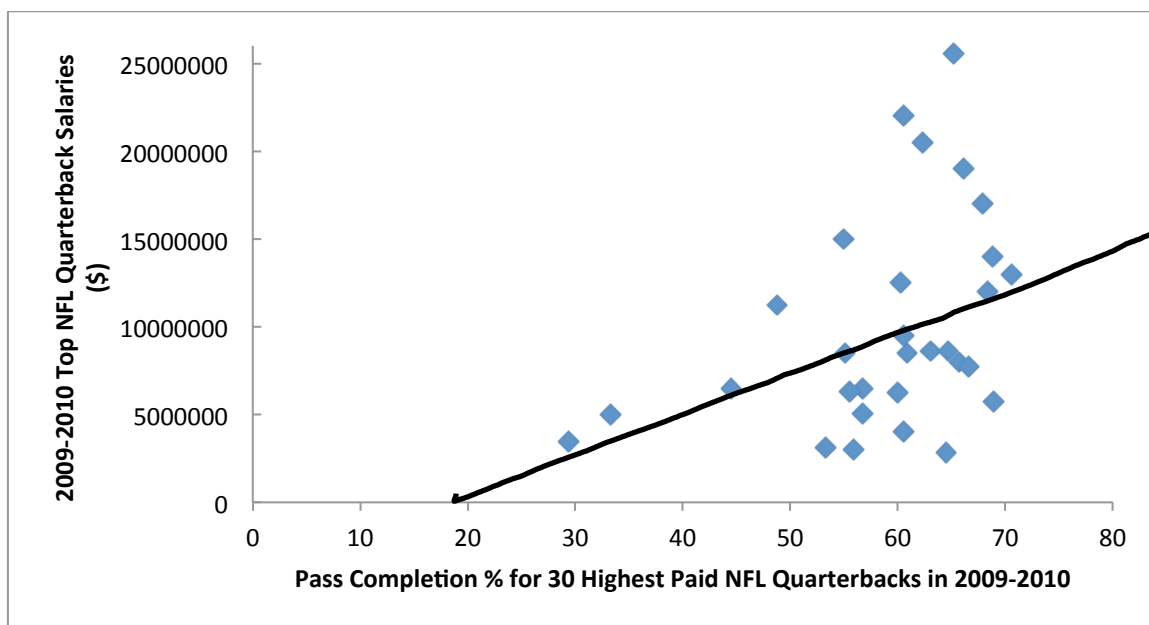
OR

Can we use the pass completion percentages of the 30 top paid NFL quarterbacks in 2009-2010 to predict their salaries?

Population of Interest – 30 top paid NFL quarterbacks in 2009-2010

Relationship of Interest – Relationship between the salaries of the top 30 NFL quarterbacks in 2009-2010 and the quarterbacks' pass completion percentages.

2. Students will enter the pass completion % data in L1 and the salary data in L2 in their graphing calculators.
3. Students should sketch the following graph found on their graphing calculator (without the regression line). On question #5, students will add the regression line.



4. **Sample Regression Equation**

$$\hat{y} = -4,443,755.852 + 244,351.3169x$$

\hat{y} = predicted salary of 2009-2010 top paid NFL quarterback in \$

x = pass completion % for top paid NFL quarterback in 2009-2010

5. **Graph of Regression Line** – See graph on #3

6. **Correlation r**

$r = 0.3940$ There is a weak to moderate positive linear relationship between the pass completion % of the 30 top paid NFL quarterbacks in 2009-2010 and their salaries.

7. Coefficient of Determination r^2

$$r^2 = 0.1553$$

15.53% of the variation in the salaries of the 30 top paid NFL quarterbacks in 2009-2010 is explained by the straight-line relationship between their pass completion % and salaries. This means that 84.47% of the variation in salaries is explained by factors other than the quarterbacks' pass completion percentages.

8. (a) A value of $r = 0.98$ would tell us that there is a strong, positive, linear association between the salaries of the 30 top paid NFL quarterbacks in 2009-2010 and the quarterbacks' pass completion percentages.
- (b) The strong association would not have been evidence that a quarterback's high pass completion percentage caused his salary to increase. Although a correlation $r = 0.98$ would indicate a strong linear relationship between a quarterback's pass completion percentage and his salary, we cannot conclude that an increase in a quarterback's salary is **caused** by his high pass completion percentage. There could be other variables that contribute to the relationship between the two variables. A strong association between two variables is not enough to draw conclusions about cause and effect.
9. The explanatory variable that seems to be the best predictor of a top paid 2009-2010 NFL quarterback's salary is the quarterback's total number of touchdowns. This decision is justified by the following:
- (1) The correlation $r = 0.5954$ for the linear relationship between the salary of a top paid NFL quarterback in 2009-2010 and the quarterback's total number of touchdowns is higher than the correlation r values for the linear relationships between the salary of a top paid NFL quarterback in 2009-2010 and the other two explanatory variables. This tells us that the linear relationship between the salary and total number of touchdowns is stronger than the linear relationships between the salary and the other two explanatory variables.
- (2) The coefficient of determination for the linear relationship between the salary of a top paid NFL quarterback in 2009-2010 and the quarterback's total number of touchdowns is $r^2 = 0.3545$. This is higher than the r^2 values for quarterback salaries versus the other two explanatory variables.
- This tells us that a larger percentage of the variation in the salaries of the 30 top paid NFL quarterbacks in 2009-2010 is explained by the straight-line relationship between their total number of touchdowns and their salaries than by the straight-line relationships between the other two explanatory variables and their salaries.

Answers to Activity Sheet for Number of Touchdowns vs. Salary

1. **Statistical Question** - Answers will vary, but should be similar to the following:

Is there an association between the 30 top NFL quarterback salaries in 2009-2010 and the quarterback's total number of touchdowns in 2009-2010?

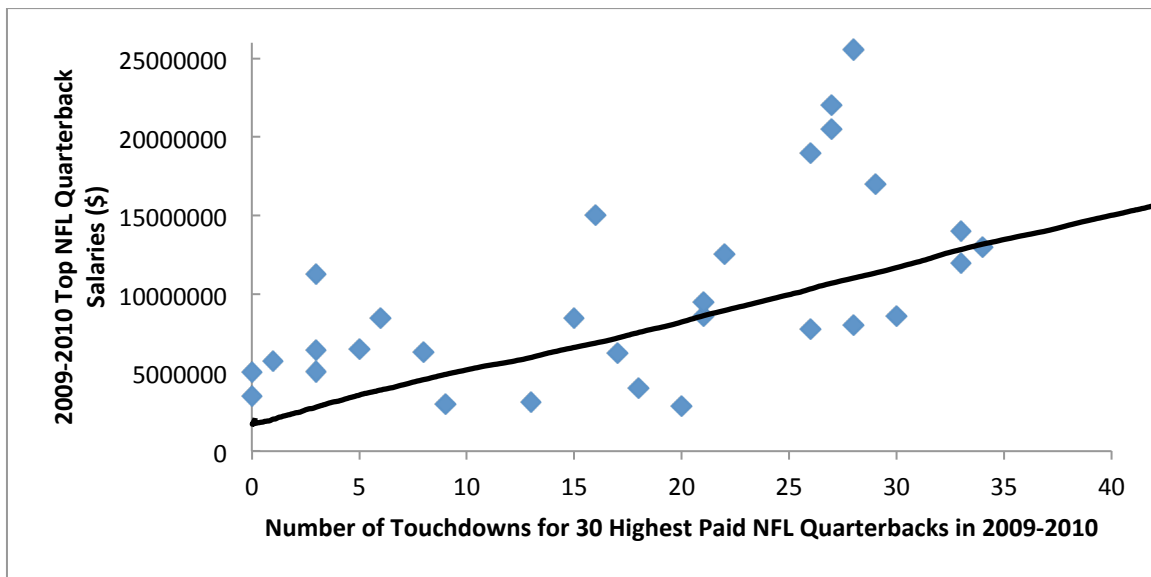
OR

Can we use the total number of touchdowns of the 30 top paid NFL quarterbacks in 2009-2010 to predict their salaries?

Population of Interest – 30 top paid NFL quarterbacks in 2009-2010

Relationship of Interest – Relationship between the salaries of the top 30 NFL quarterback in 2009-2010 and their total number of touchdowns.

2. Students will enter the pass completion % data in L1 and the salary data in L2 in their graphing calculators.
3. Students should sketch the following graph found on their graphing calculator (without the regression line). Students will add the regression line in question #5.



4. **Sample Regression Equation**

$$\hat{y} = 4,393,649.841 + 320,510.258x$$

\hat{y} = predicted salary of 2009-2010 top paid NFL quarterback in \$

x = total number of touchdowns in 2009-2010 of top paid NFL quarterback

5. **Graph of Regression Line** – see graph on #3

6. **Correlation r**

$r = 0.5954$ There is a moderate positive linear relationship between the total number of touchdowns of the 30 top paid NFL quarterbacks in 2009-2010 and their salaries.

7. **Coefficient of Determination r^2**

$$r^2 = 0.3545$$

35.45% of the variation in the salaries of the 30 top paid NFL quarterbacks in 2009-2010 is explained by the straight-line relationship between their total number of touchdowns and their salaries. This means that 64.55% of the variation in salaries is explained by factors other than their total number of touchdowns.

8. (a) A value of $r = 0.98$ would tell us that there is a strong, positive, linear association between the salaries of the 30 top paid NFL quarterbacks in 2009-2010 and the quarterbacks' total number of touchdowns.

(b) The strong association would not have been evidence that a quarterback's large number of touchdowns caused his salary to increase. Although a correlation $r = 0.98$ would indicate a strong linear relationship between a quarterback's number of touchdowns and his salary, we cannot conclude that an increase in a quarterback's salary is **caused** by his large number of

touchdowns. There could be other variables that contribute to the relationship between the two variables. A strong association between two variables is not enough to draw conclusions about cause and effect.

9. The explanatory variable that seems to be the best predictor of a top paid 2009-2010 NFL quarterback's salary is the quarterback's total number of touchdowns. This decision is justified by the following:
- (1) The correlation $r = 0.5954$ for the linear relationship between the salary of a top paid NFL quarterback in 2009-2010 and the quarterback's total number of touchdowns is higher than the correlation r values for the linear relationships between the salary of a top paid NFL quarterback in 2009-2010 and the other two explanatory variables. This tells us that the linear relationship between the salary and total number of touchdowns is stronger than the linear relationships between the salary and the other two explanatory variables.
 - (2) The coefficient of determination for the linear relationship between the salary of a top paid NFL quarterback in 2009-2010 and the quarterback's total number of touchdowns is $r^2 = 0.3545$. This is higher than the r^2 values for quarterback salaries versus the other two explanatory variables.
This tells us that a larger percentage of the variation in the salaries of the 30 top paid NFL quarterbacks in 2009-2010 is explained by the straight-line relationship between their total number of touchdowns and their salaries than by the straight-line relationships between the other two explanatory variables and their salaries.

Answers to Activity Sheet for Average Number of Yards per Game vs. Salary

1. **Statistical Question** - Answers will vary, but should be similar to the following:

Is there an association between the 30 top NFL quarterback salaries in 2009-2010 and the quarterbacks' average number of yards per game?

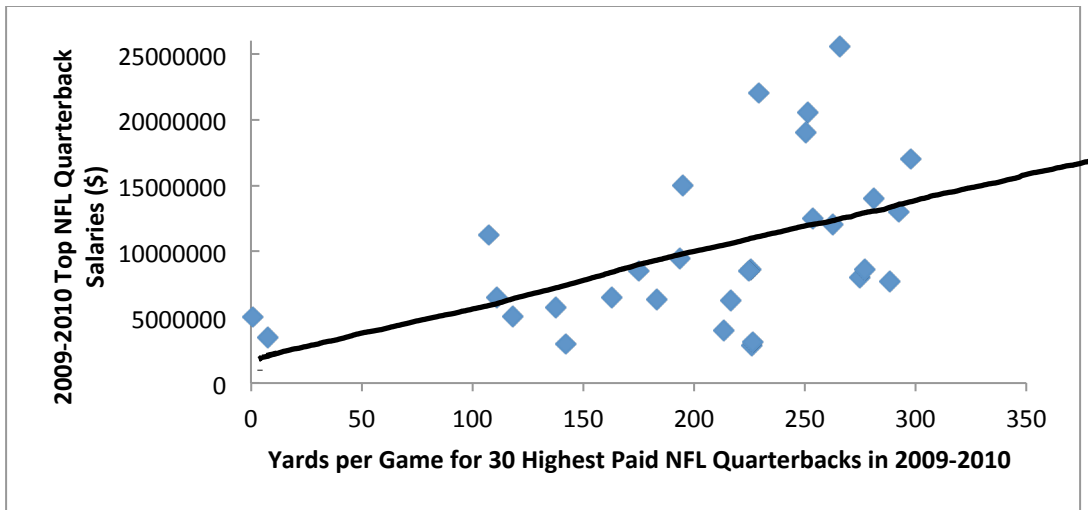
OR

Can we use the average number of yards per game of the 30 top paid NFL quarterbacks in 2009 - 2010 to predict their salaries?

Population of Interest - 30 top paid NFL quarterbacks in 2009-2010

Relationship of Interest – Relationship between the salaries of the top 30 NFL quarterbacks in 2009 - 2010 and the quarterbacks' average number of yards per game.

2. Students will enter the yards per game data in L1 and the salary data in L2 in their graphing calculators.
3. Students should sketch the following graph found on their graphing calculator (without the regression line). On question #5, students will add the regression line.



4. **Sample Regression Equation**

$$\hat{y} = 2,242,948.153 + 38,047.530x$$

\hat{y} = predicted salary of 2009-2010 top paid NFL quarterback in \$

x = yards per game in 2009-2010 for top paid NFL quarterback

5. **Graph of Regression Line** – See graph on #3

6. **Correlation r**

$r = 0.4886$ There is a moderate positive linear relationship between the yards per game of the 30 top paid NFL quarterbacks in 2009-2010 and their salaries.

7. **Coefficient of Determination r^2**

$$r^2 = 0.2387$$

23.87% of the variation in the salaries of the 30 top paid NFL quarterbacks in 2009-2010 is explained by the straight-line relationship between their yards per game and their salaries. This means that 76.13% of the variation in salaries is explained by factors other than their yards per game.

8. (a) A value of $r = 0.98$ would tell us that there is a strong, positive, linear association between the salaries of the 30 top paid NFL quarterbacks in 2009-2010 and the quarterbacks' average number of yards per game.

(b) The strong association would not have been evidence that a quarterback's high average number of yards per game caused his salary to increase. Although a correlation $r = 0.98$ would indicate a strong linear relationship between a quarterback's average number of yards per game and his salary, we cannot conclude that an increase in a quarterback's salary is **caused** by his high average number of yards per game. There could be other variables that contribute to the relationship between the two variables. A strong association between two variables is not enough to draw conclusions about cause and effect.

9. The explanatory variable that seems to be the best predictor of a top paid 2009-2010 NFL quarterback's salary is the quarterback's total number of touchdowns. This decision is justified by the following:

- (1) The correlation $r = 0.5954$ for the linear relationship between the salary of a top paid NFL quarterback in 2009-2010 and the quarterback's total number of touchdowns is higher than the correlation r values for the linear relationships between the salary of a top paid NFL quarterback in 2009-2010 and the other two explanatory variables. This tells us that the linear relationship between the salary and total number of touchdowns is stronger than the linear relationships between the salary and the other two explanatory variables.
- (2) The coefficient of determination for the linear relationship between the salary of a top paid NFL quarterback in 2009-2010 and the quarterback's total number of touchdowns is $r^2 = 0.3545$. This is higher than the r^2 values for quarterback salaries versus the other two explanatory variables.
This tells us that a larger percentage of the variation in the salaries of the 30 top paid NFL quarterbacks in 2009-2010 is explained by the straight-line relationship between their total number of touchdowns and their salaries than by the straight-line relationships between the other two explanatory variables and their salaries.