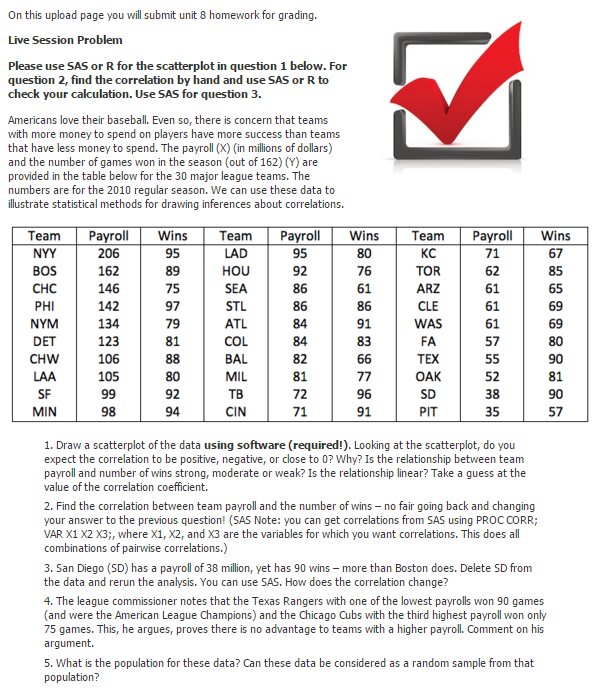
UNIT 8 HW

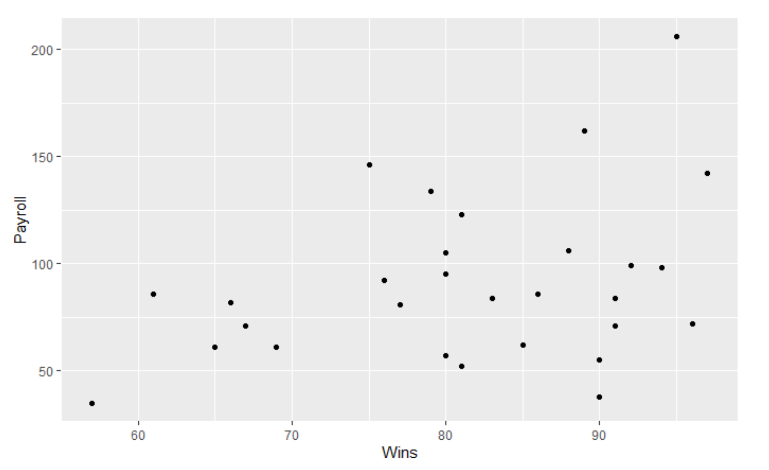
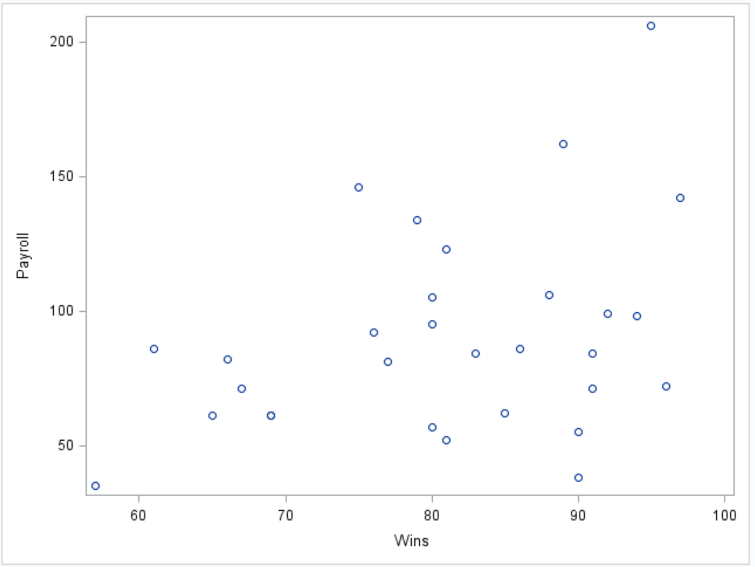


Americans love their baseball. Even so, there is a concern that teams with more money to spend on players have more success than teams that have less money to spend. The payroll (X) (in millions of dollars) and the number of games won in the season (out of 162) (Y) are provided in the table below for all of the 30 major league teams. The numbers are from the 2010 regular season. We can use these data to illustrate statistical methods for drawing inferences about correlations.



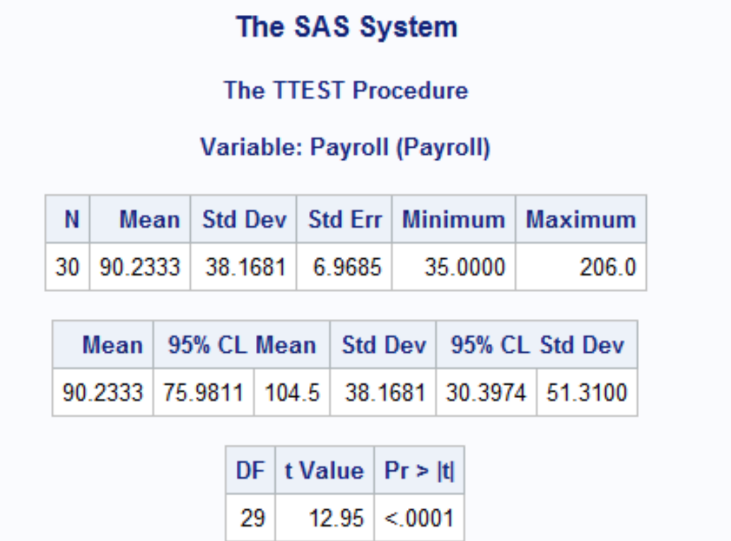
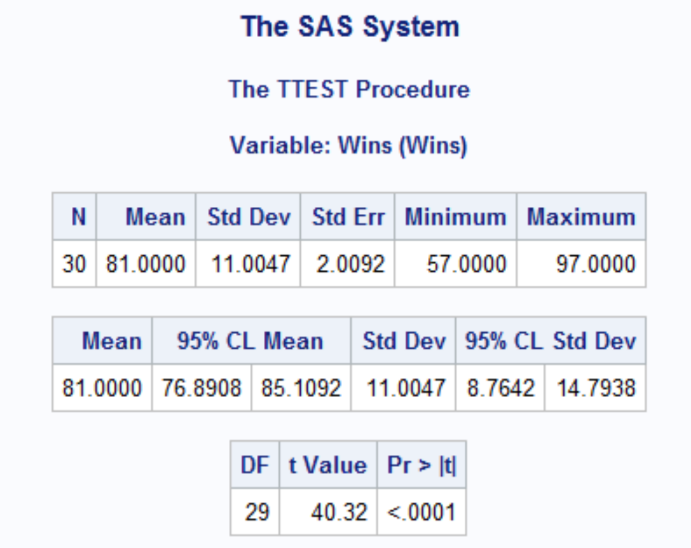
1. Provide a scatterplot of the data using both SAS and R. Looking at the scatterplot, do you expect the correlation to be positive, negative, or close to 0? Why? Is the relationship between team payroll and number of wins strong, moderate, or weak? Is the relationship linear? Take a guess of the value of the correlation coefficient.

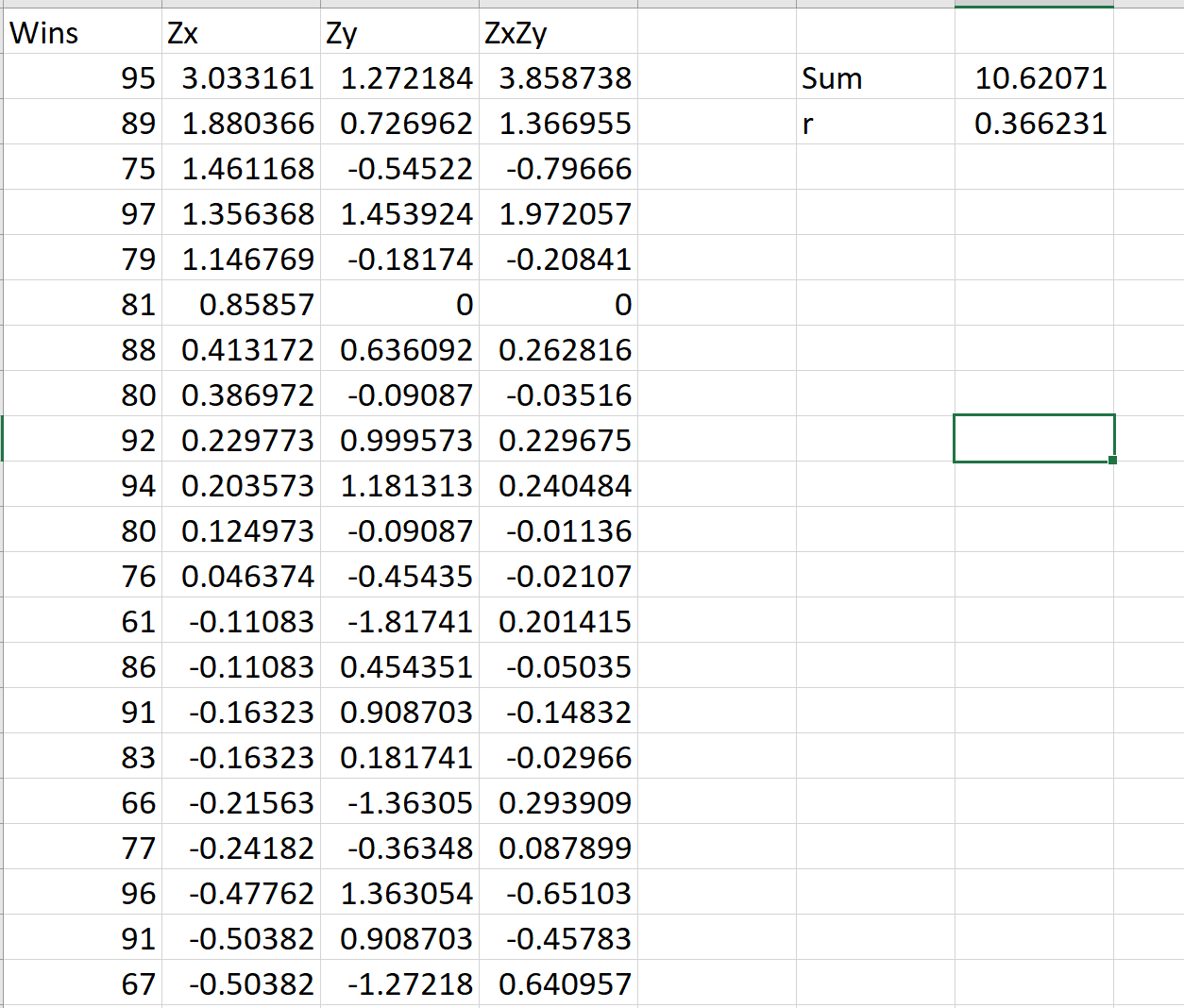
**SAS & R Scatterplots**



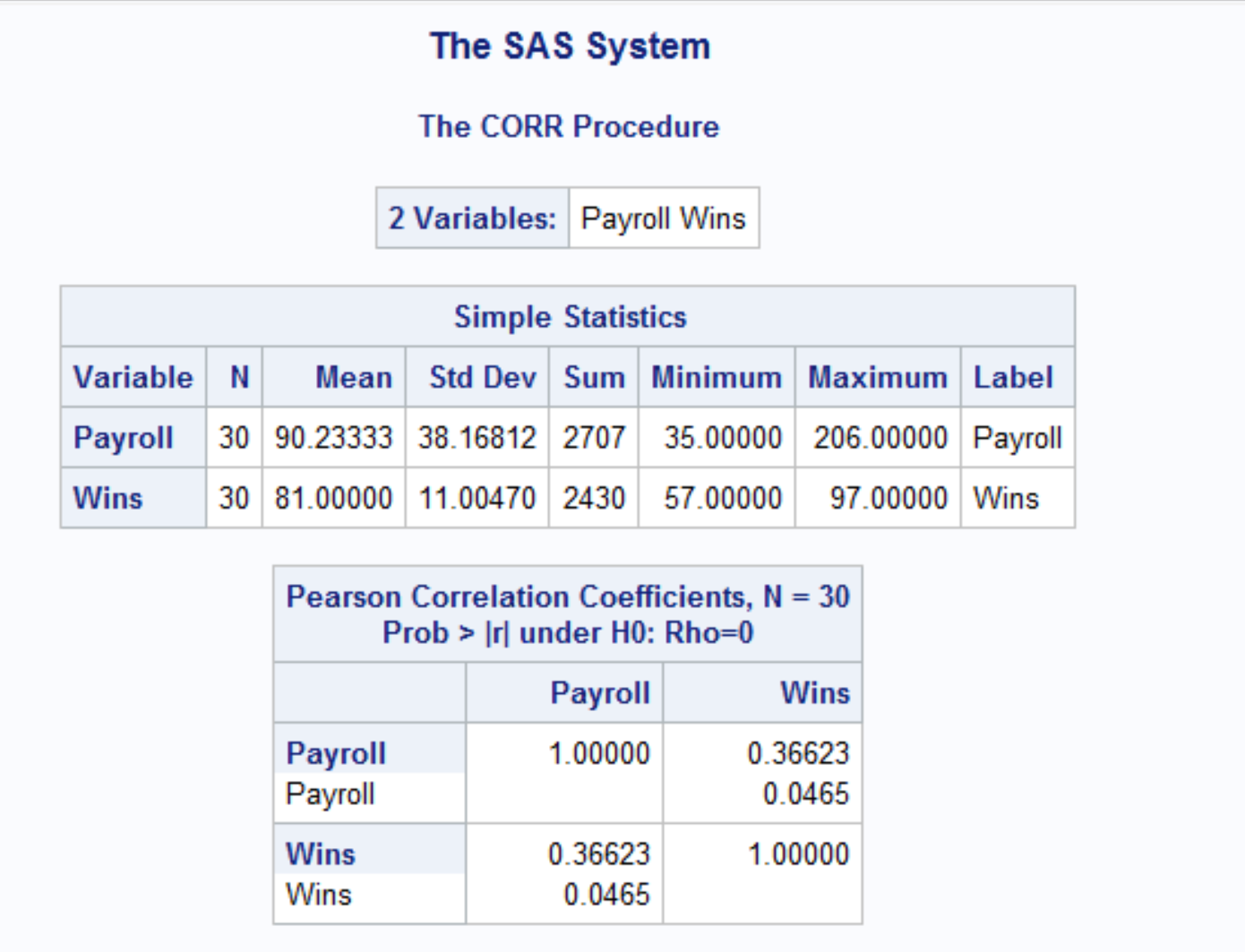
The correlation will be positive since you can make out weak line. The wins tend to be higher with more on payroll. But it is not a strong correlation and the relationship seems to be linear. My estimate for correlation coefficient to be 0.5 for the payroll and wins.

1. Find the correlation between team payroll and the number of wins. (No fair going back and changing your answer to the previous question!) Do this manually with the means and standard deviations of payroll & number of wins. Confirm your answer in R and SAS.

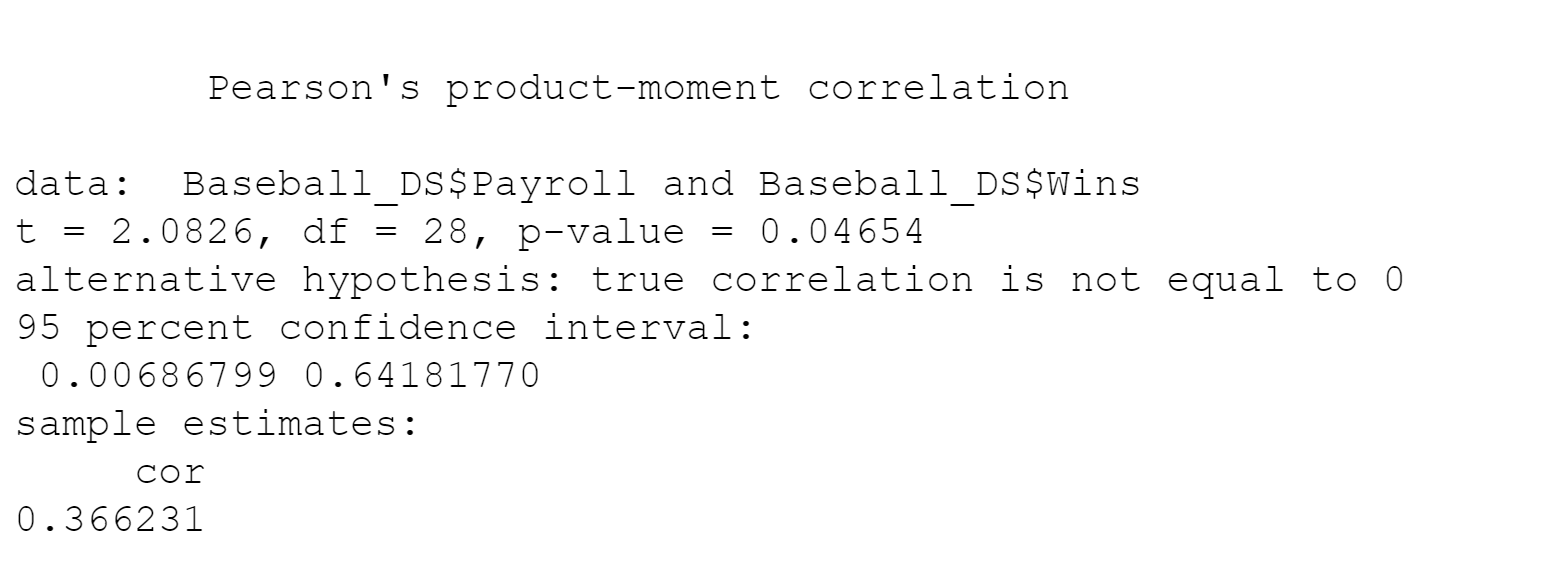




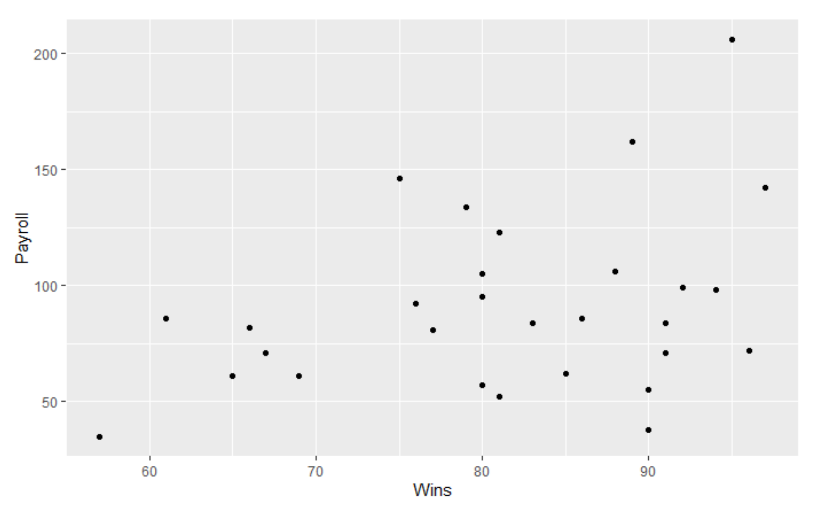
Verify in SAS the correlation:

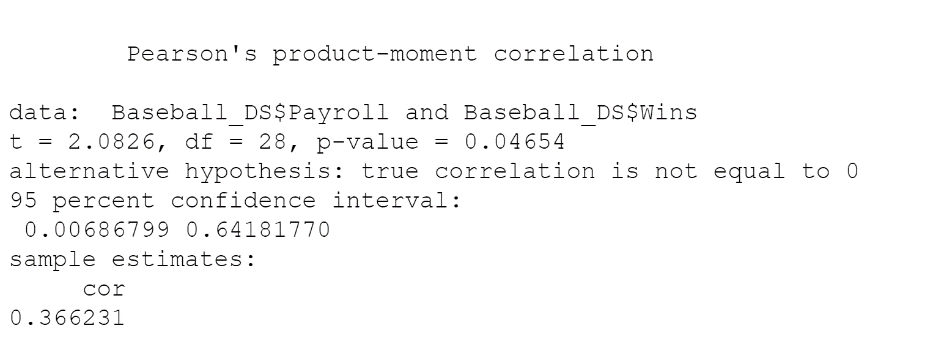


Verify in R the correlation:

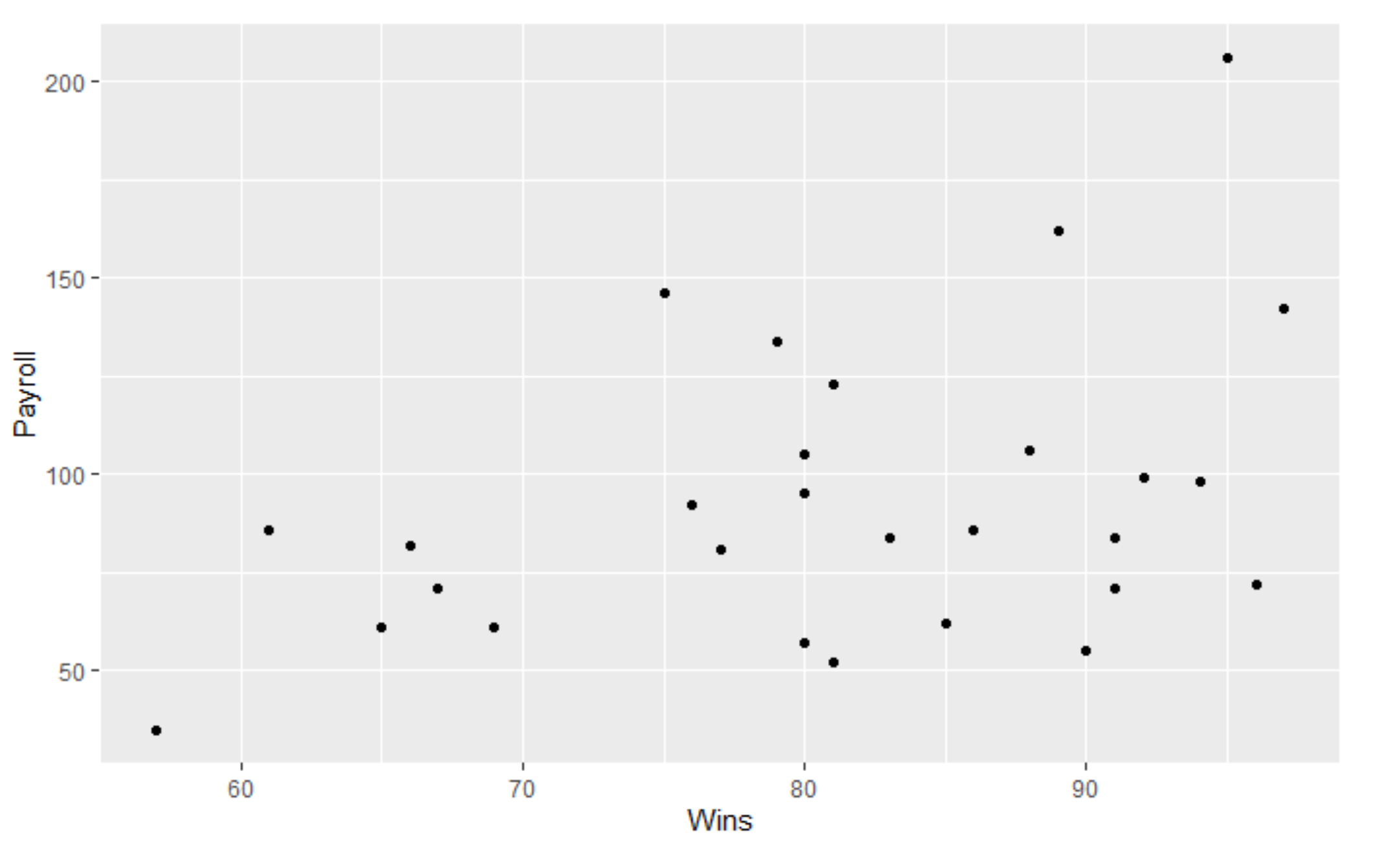
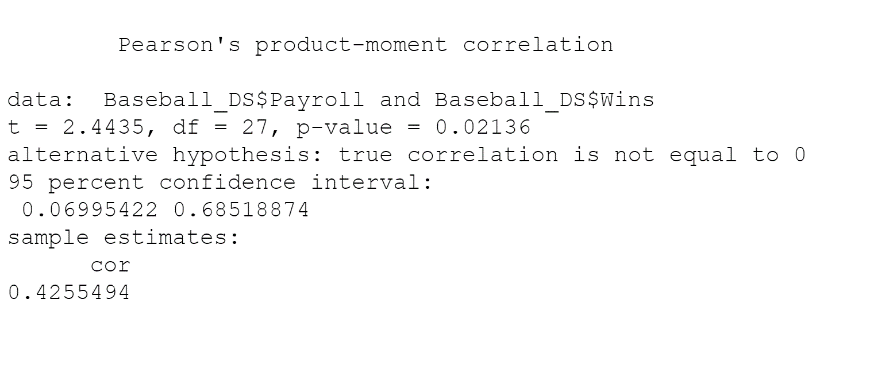


1. Manually calculate the test statistic for the correlation in question 2 and report the probability of observing the value you calculate.
2. San Diego (SD) has a payroll of $38 million, yet SD has 90 wins – more than Boston does (with a payroll of $162 million). Delete SD from the data and rerun the analysis (scatter plot and correlation value). How does the correlation change? You may use your preference here, R or SAS.

Original run:



Rerun after SD deletion:



The correlation got stronger since it went from 0.36 to 0.42. Still the correlation between payroll and wins is not strong.

1. The league commissioner notes that the Texas Rangers (TEX), with one of the lowest payrolls, won 90 games (and were the American League Champions) and the Chicago Cubs (CHC), with the third highest payroll, won only 75 games. He argues that this proves that there is no advantage to teams with a higher payroll. Comment on his argument.

Higher payroll will not always produce the highest winning team. You still need to insure that your players in the team are good. But practical significance the higher the payroll the more valuable your team will be since the top talent commands higher salaries. We would have to review other factors such as injuries, weather etc to understand the real correlation.

1. What is the population for these data? Can these data be considered a random sample from that population?

The population for this data is the baseball teams. The data cannot be considered random sample since you have covered all the baseball teams in the major league.