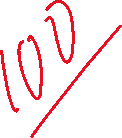
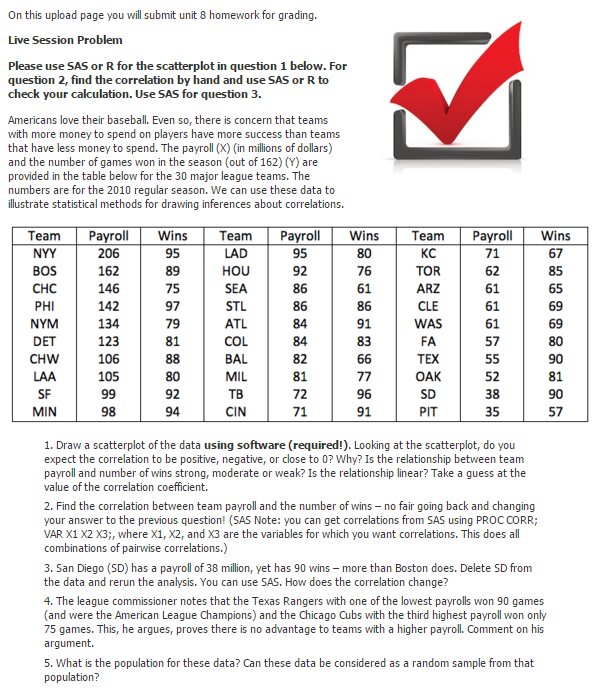
UNIT 8 HW



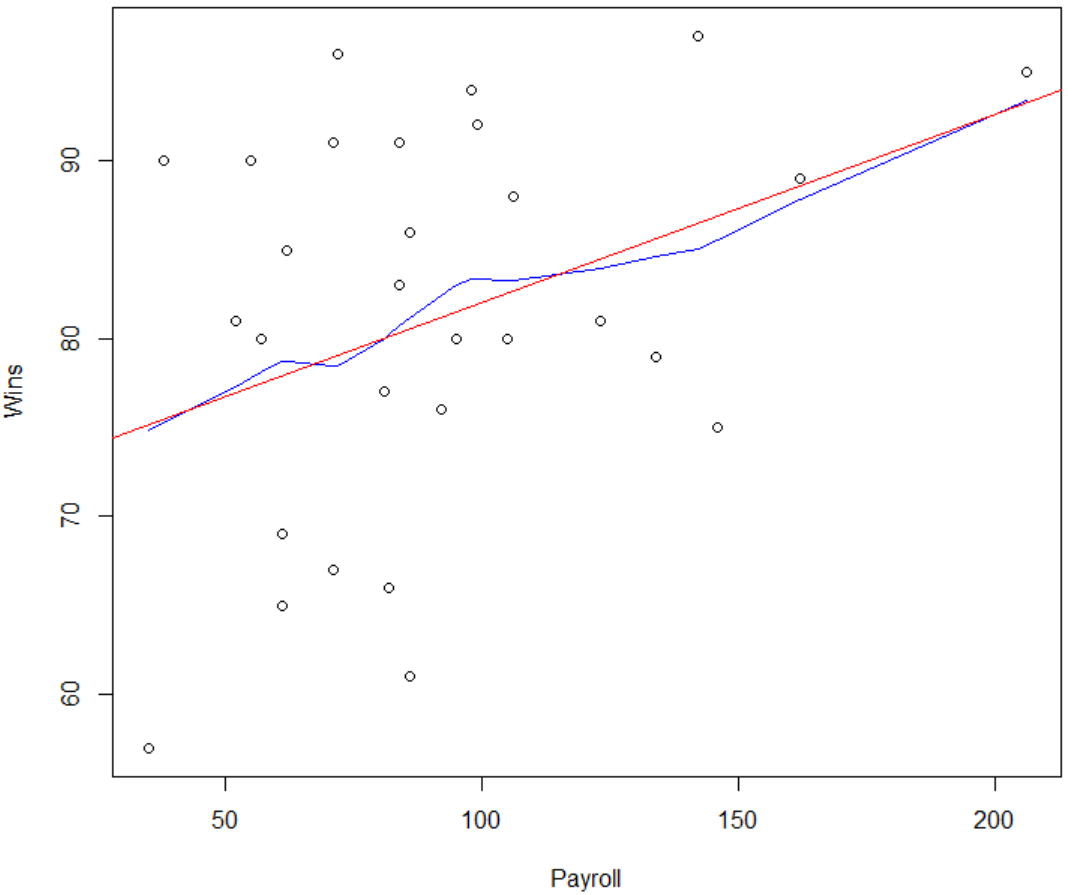
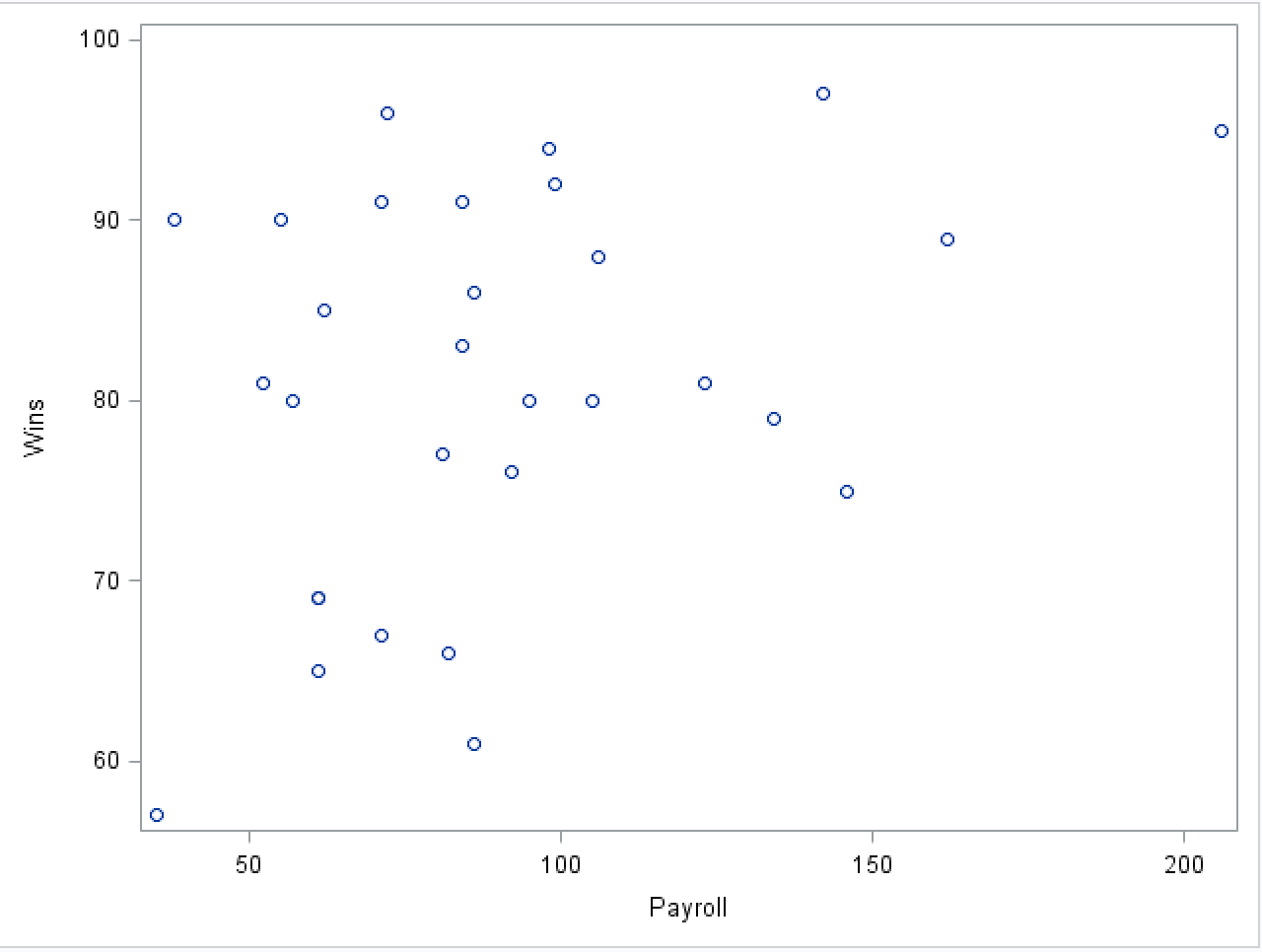


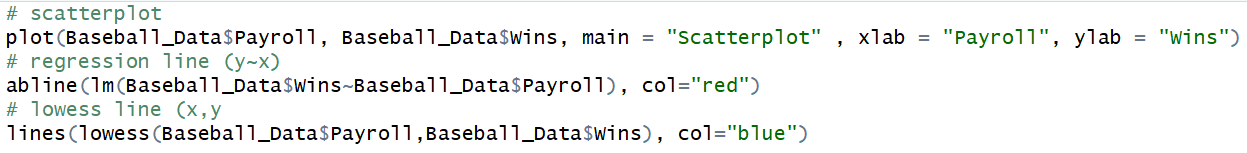
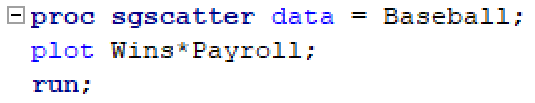
Americans love their baseball. Even so, there is a concern that teams with more money to spend on players have more success than teams that have less money to spend. The payroll (X) (in millions of dollars) and the number of games won in the season (out of 162) (Y) are provided in the table below for 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.

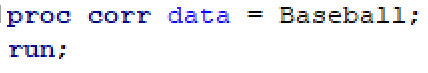
The correlation appears to be positive. As payroll increases, you can see the number of wins also increases. However, the strength of the relationship is weak at best since there is very high variance in the data. I would guess the coefficient to be around 25%.

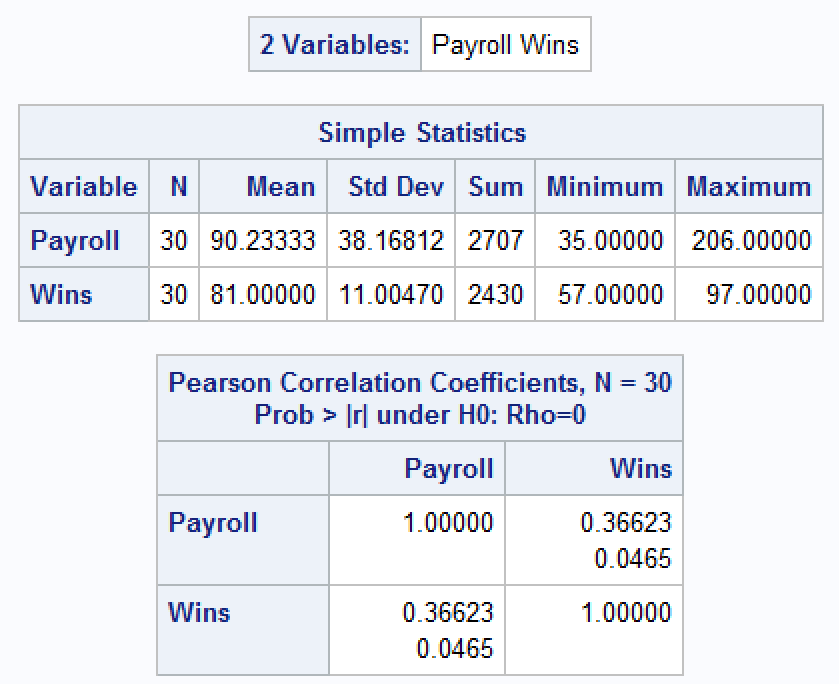
 

1. Find the correlation between team payroll and the number of wins. (No fair going back and changing your answer to the previous question!) You should do this in both R and SAS.

SAS and R both confirm the coefficient of the data (r) = 0.366.

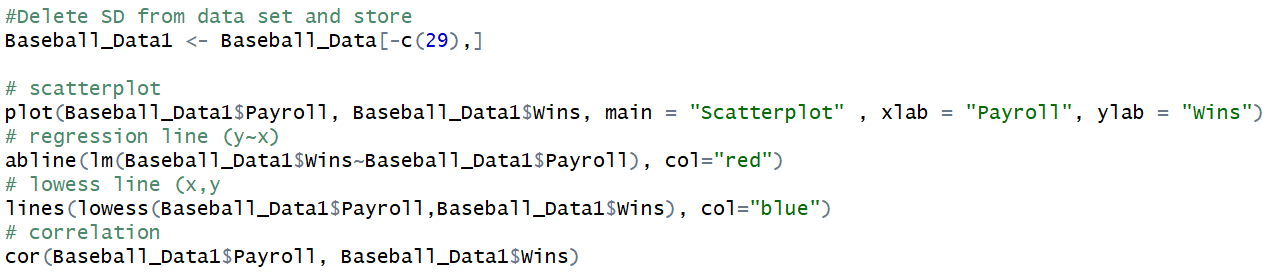


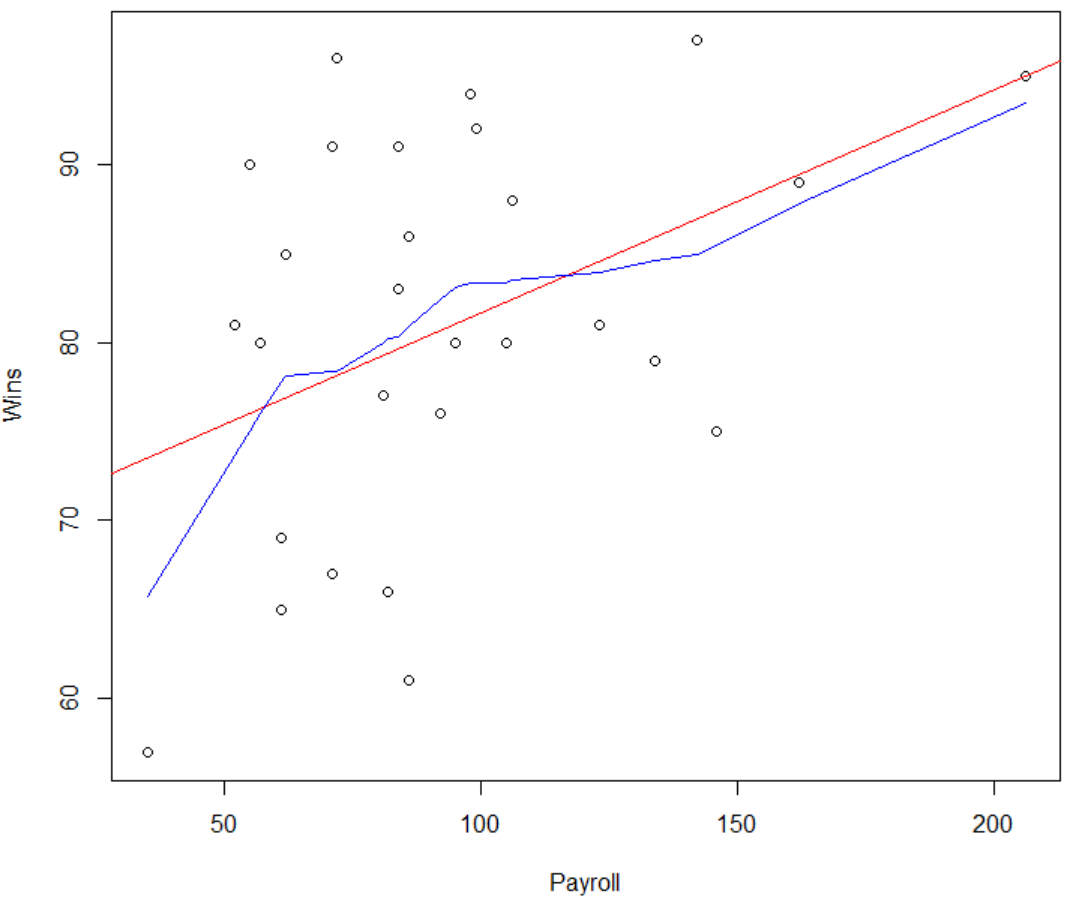




1. San Diego (SD) has a payroll of $38 million, yet SD has 90 wins – more than Boston does (with a payroll of $162 million). Delete SD from the data and rerun the analysis (scatter plot and correlation value). How does the correlation change? You may use your preference here, R or SAS.

The strength of the correlation actually increases from 0.366 to 0.426. Considering San Diego was one of the teams with a low payroll and high wins, by removing it from the data will make the linear model closer to 1.





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.

While the commissioner’s observations are correct, they are overgeneralizing the data. Sure, the Texas Rangers won more games with a lower payroll. Yet, there are so many other variables that would have to be constant in order to truly make that observation. Considering the MLB is a very competitive organization, the extraneous variables are far from constant. The skill level of the front office and coaching staff probably have the biggest impact on payroll and games won than any other variables. These are the parties that make decisions regarding salary and plan for future games. Players are constantly seeking higher contracts and are willing to go to other teams who are willing to pay it. A contract year and strong performance could result in an inflated contract, in which performance may begin to decrease. Front office will argue that the skill trade off translates into leadership in the locker room and so on. However, that is just an example of variables that are not being analyzed by this data set. As a result, while the observation of the commissioner is true. It is simply too early to make that observation. What we know based on this data set is that payroll only explains 36% of the variance relating to games. In other words, having the most expensive players on the roster only constitutes for 36% of what it takes to get the win. At the end of the day, there is 64% of unexplained variance that needs to be accounted for through additional studies.

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

There are 30 teams in the Major League Baseball. Considering there are 30 “samples” in this data set, the data set is comprised of the entire population of teams in the US. Although this may not be classified as a “random sample,” this can be deemed proper is certain situations in which the population may be too small (such of this one) and/or the population has uncommon characteristics that make it stand out. In this instance, MLB teams have unlimited salary cap (which means they can pay as much as they can afford), which can inflate their numbers if it were to be compared to teams from other countries. As a result, any studies trying to address this question at an international level would need to address the fact that it is possible for every single MLB team to be an “outlier” in the data.