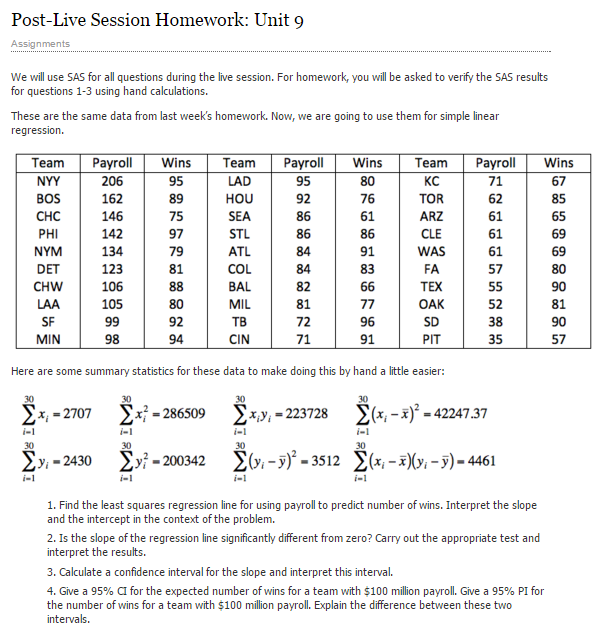
UNIT 9 HW

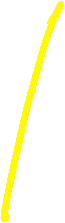
These are the same data from last week’s HW. Now, we are going to use them for simple linear regression.





1) a.

i. Find the least squares regression line using payroll to predict the number of wins. Interpret the slope and the intercept in the context of the problem. Show your work in finding the slope and intercept. You will need the above calculations. Do this by hand or using a basic calculator, but **NOT** by uploading the data into software. There are several equivalent formulations for the elements of the least squares regression line ( and ). Find one that utilizes the series (sums) above.



ii. Interpret the slope **AND** the intercept in the context of the problem.



The y-intercept of the model indicates that should a team be paid the minimum allotted by the MLB, the mean games won would be 71.47205. The slope indicates that for every additional $1M on top of this minimum, the mean additional games won will be 0.105592 or for every $10M more spent in payroll, the mean additional games won will be 1.005592 more.

b. Is the slope (only concerned with the slope here) of the regression line significantly different from zero? Carry out a 6-step hypothesis test to address this question. Use the above calculations to find the relevant statistics for this test. You will need to use SAS, R, the internet, a calculator, or integration to find the p-value and critical value, but do NOT upload the data to software. (One of the first 4 choices is suggested. ☺) Use α = 0.05.

1.

2. Critical Value: ± 2.048

3.

4. p value = 0.0465



5. Reject H0



6. There is sufficient evidence at the alpha = 0.05 level of significance (p-value = 0.0028) to suggest that the data are linearly correlated.

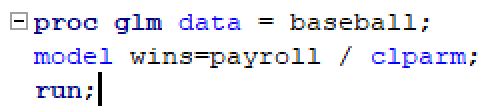
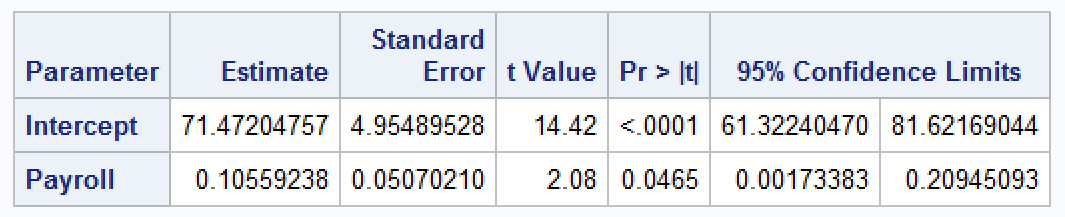
c.

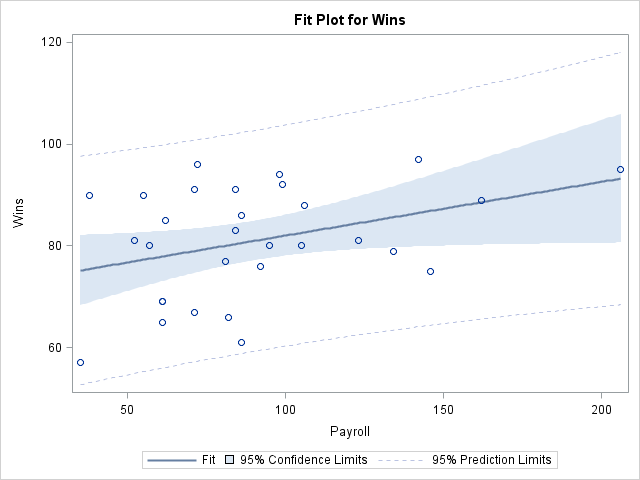
i. **BY HAND (**or basic calculator), calculate a 95% confidence interval for the slope. You should already have the pieces of the confidence interval (point estimate, multiplier, and standard error) from part 1b.

0.20943, 0.001754]

ii. Interpret the interval.

We are 95% confident that when the payroll is increased by $1M, the mean games won increases between 0.20943 and 0.001754.

d. Verify your results (parameter estimates, test statistic for the hypothesis test of whether the slope equals zero, p-value for this same hypothesis test, and confidence interval for the slope) with SAS. Paste your code and relevant output below. Note what is the same or different. 

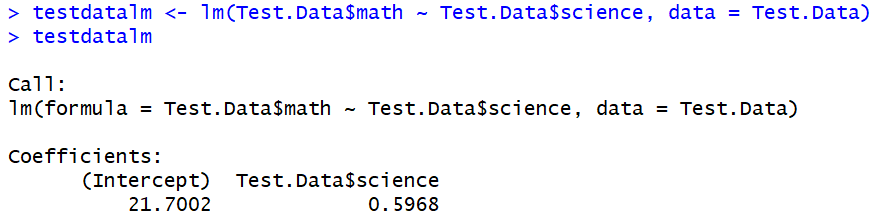


In this instance, the numbers match for the most part. Except, there appears to be some rounding error once we get into the thousandth values. The intercept, slope, SEs, t-statistic, and p-value all match. Even the confidence intervals match as well. However, there is slight difference as mentioned prior.

2)

a.

i. Find the least squares regression line to assess the relationship between the math and the science score for the Test Data. We would like to be able to estimate a change in the mean math score for a one point change in the mean science score. (This should help identify the response and the independent variables.) Write your regression equation and paste your code and relevant output below. You should obtain the test statistics and other relevant statistics from R.



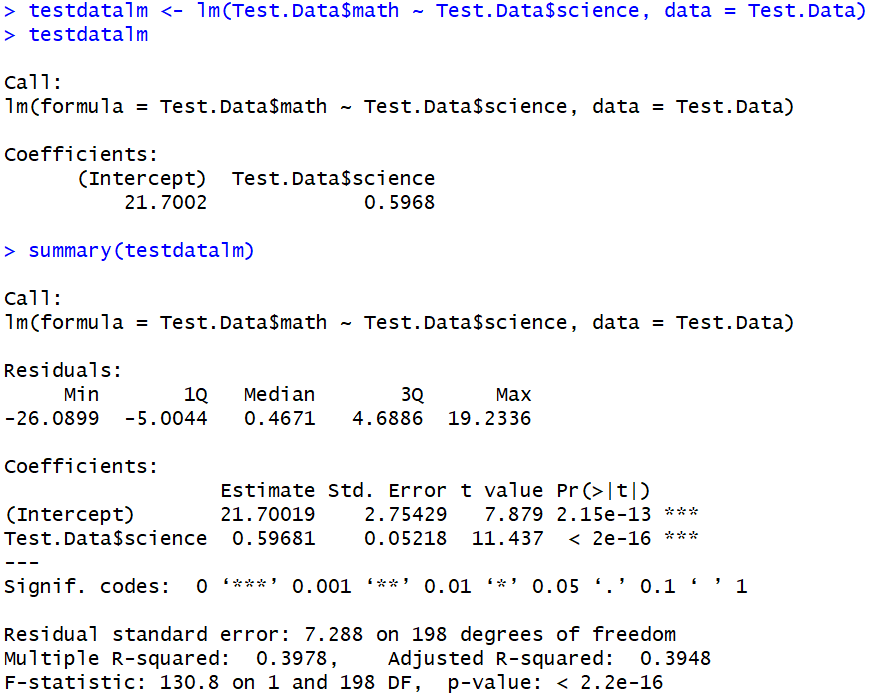
ii. Interpret the slope and the intercept in the context of the math and science scores.

The intercept value indicates that in the event the student gets the lowest score possible on the science test (26 in this data set), the mean math test score is going to be 21.7002. Yet, for every point increased in the science test score, the mean math score will increase by 0.5968 points.

b. Are the slope ***and intercept*** of the regression line significantly different than zero? Carry out a 6-step hypothesis test **for each** regression parameter to address this question (two different hypothesis tests). You should obtain the test statistics and other relevant statistics from R. Paste your code and any relevant output below. Use alpha = 0.01.



|  |  |  |  |
| --- | --- | --- | --- |
| 1. |  | 1. |  |
| 2. | Critical Value: ± 2.763 | 2. | Critical Value: ± 2.763 |
| 3. |  | 3. |  |
| 4. | p-value = <0.001 | 4. | p-value = <0.001 |
| 5. | Reject null | 5. | Reject null |
| 6. | There is sufficient evidence at the alpha = 0.01 level of significance (p-value < 0.001) to suggest that the data are linearly correlated. | 6. | There is sufficient evidence at the alpha = 0.01 level of significance (p-value < 0.001) to suggest that the y-intercept is not equal to zero. |





c.

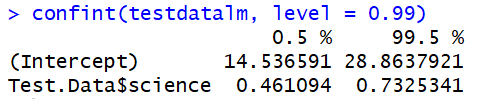
**i. BY HAND**, calculate 99% confidence intervals for the slope and intercept (**two** separate confidence intervals). You may use point estimates, multipliers, and standard errors found from software, but put these pieces together to form confidence intervals by hand (or basic calculator).

ii. Interpret these intervals.

We are 99% confidence that for every additional point in the science test, the mean math test score will increase between 0.46 and 0.73 points.

We are 99% confident that the intercept is between 14.5 and 28.9 points.

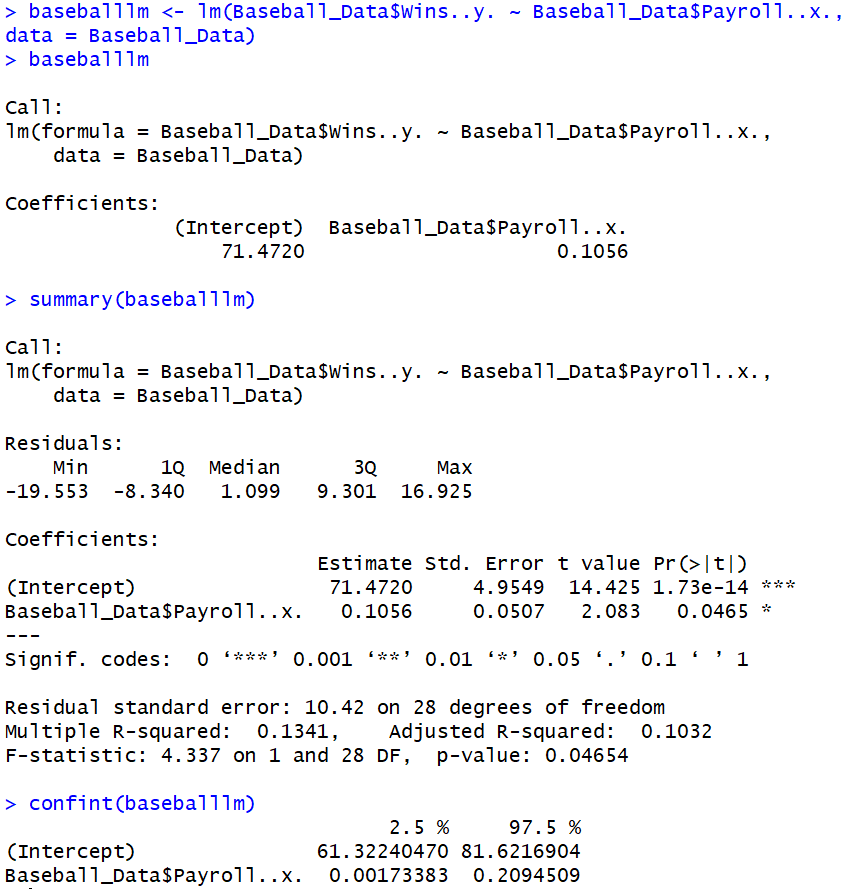
d. Verify your confidence intervals (for and ) with R and paste your code and relevant output below.



BONUS:

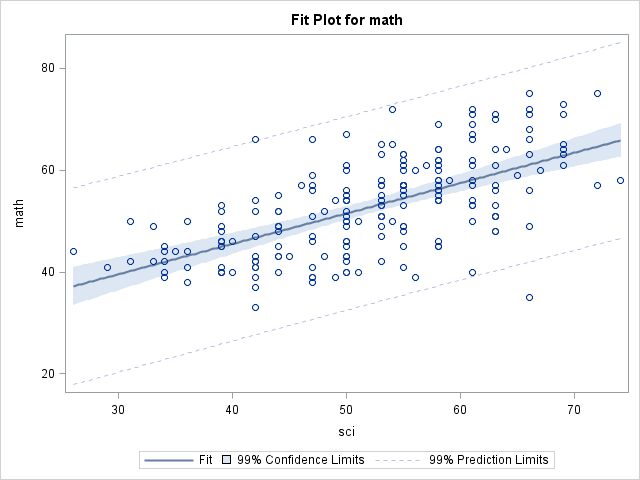
3) Repeat 1(d) using R.

As mentioned in the prior answer, the results are pretty identical with some slight variance in the thousandth decimal values.

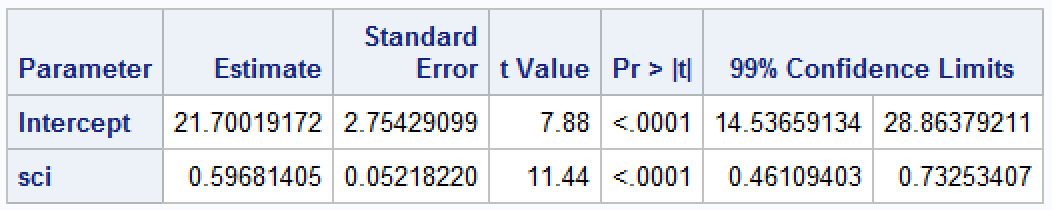


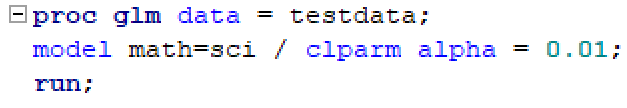


4) Repeat 2(a)(i) and 2(d) using SAS.









5) We will cover this in Unit 10 ….

With reference to the baseball data … we will learn how to do the following next week.

1. Give a 95% CI (confidence interval) for the expected number of wins for a team with $100 million payroll. Use SAS or R.
2. Give a 95% PI (prediction interval) for the number of wins for a team with $100 million payroll. Use SAS or R.
3. Explain the difference between these two intervals.