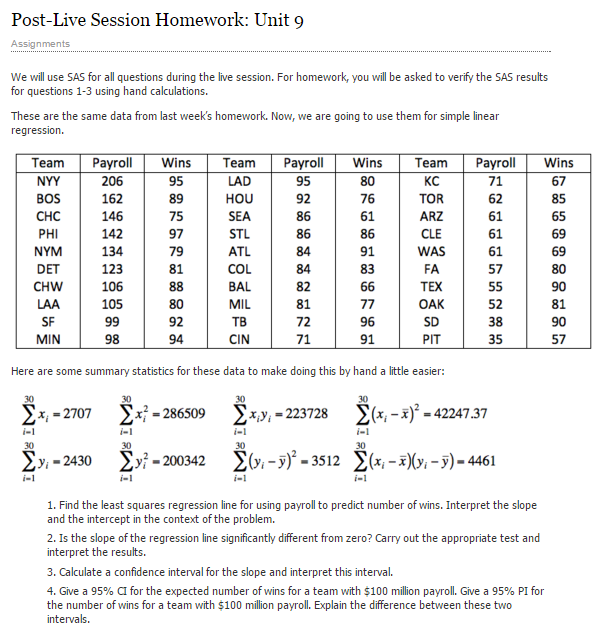
UNIT 9 HW

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



1a. Find the least squares regression line for 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.

The interpretation of the above equation is that if you somehow had a professional team that cost 0 dollars you would expect 71.47 wins. Then for each additional million dollars spent, you should expect the number of wins by the team to increase by 0.1056.

b. Is the slope (just the slope) of the regression line significantly different than 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 or the internet or a calculator or integration to find the pvalue (I suggest 1 of the first 3. ☺)

**Problem:** Is the slope of the regression line significantly different from 0?

**Hypothesis:**

H0: The slope of the regression line equals 0

HA: The slope of the regression line does not equal 0

**Find Critical Values**

The critical value is 2.048 (Assuming alpha of 0.05, a two tailed test, with 28 degrees of freedom)

|  |
| --- |
| data work.critvalue;  critVal = quantile("T", 0.975, 28); /\*distribution, area, degree of freedom\*/  run; |

**Analysis**

T-score = 2.07 (associated p-value of 0.0465)

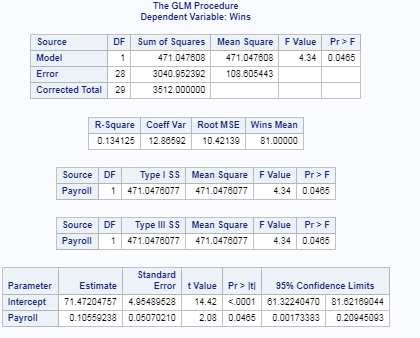
**Conclusion**

Reject the null hypothesis. There is evidence to suggest that the slope of the regression line is statistically different from 0.

c. Calculate a 95% confidence interval for the slope and interpret the interval.

d. Verify your results with both SAS and R. Paste your code below.

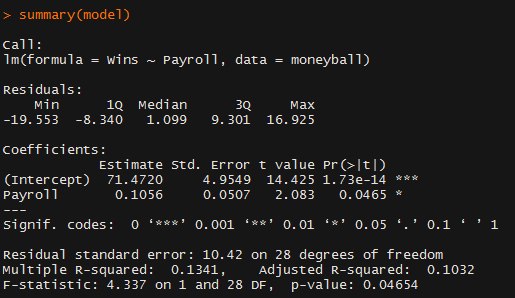
**SAS Results**



**SAS Code**

|  |
| --- |
| proc glm data=work.moneyball;  model wins = payroll / solution clparm;  run; |

**R Results**



**R Code**

|  |
| --- |
| moneyball = read.csv("moneyball\_raw.csv")  model = lm(Wins~Payroll, data=moneyball)  summary(model) |

2a. Find the least squares regression line to assess the relationship between the math and the science score. 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.) Interpret the slope and the intercept in the context of the math and science scores. You should obtain the test statistics and other relevant statistics from SAS or R. Write your regression equation and Paste your code below (SAS OR R).

**Answer**

The above equation means that for every 1 point increase in the science score that there is a 0.596 increase in the estimated math score. Additionally, if the Science score was 0 then the estimated math score is 21.7.

**SAS Code**

|  |
| --- |
| proc glm data=work.testscores;  model math = science / solution clparm;  run; |

b. Is 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. You should obtain the test statistics and other relevant statistics from SAS or R. Paste your code below. Formalize the analysis in a 6 step format.

**1. Hypothesis:**

H0a:

HAa:

H0b:

HAa:

**2. Find the Critical Value**

The critical value used in these tests is 1.972

SAS Code used to find the critical value:

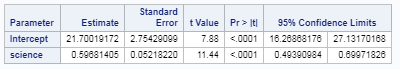
|  |
| --- |
| data work.critvalue2;  critVal = quantile("T", 0.975, 198); /\*distribution, area, degree of freedom\*/  run; |

**3 and 4. Find the T Statistic and Associated P values**

Intercept t = 7.88 with an associated p-value less than 0.0001

Slope t = 11.44 with an associated p-value less than 0.0001

See SAS output below



Code Used to find these values:

|  |
| --- |
| proc glm data=work.testscores;  model math = science / solution clparm;  run; |

**5. Results**

Based on the t scores and associated p-values the null hypothesis can be rejected for both the intercept and the slope.

**6. Conclusion**

There is sufficient evidence at an alpha of 0.05 level of significance that the intercept of the regression line does not equal zero (t-score of 7.88 with an associated p-value of less than 0.0001). Similarly, there is sufficient evidence to suggest that the slope of the regression does not equal zero (t-score of 11.44 with an associated p-value less than 0.0001).

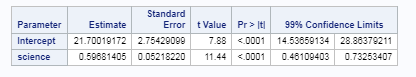
c. Calculate a 99% confidence interval for the slope and intercept and interpret these intervals (by hand).

**SAS Code used to get the t value**

|  |
| --- |
| data work.critvalue3;  critVal = quantile("T", 0.995, 198);  run; |

d. Verify your confidence interval with SAS or R and paste your code below.

My confidence intervals are nearly the same as those generated by SAS. See SAS output and code below.



|  |
| --- |
| proc glm data=work.testscores alpha=0.01;  model math = science / solution clparm;  run; |

BONUS: +5 We will cover this in Unit 10 ….

With reference to the baseball data …

Give a 95% CI for the expected number of wins for a team with $100 million payroll. Give a 95% PI for the number of wins for a team with $100 million payroll. Explain the difference between these two intervals.