ANOVA  
Which division command I should choose for continuation the sports career if I am guided only by salaries dispersion/variability in a division and I want a division where salaries least differ from the player to the player?  
title 'Baseball Data';

proc contents data=sashelp.Baseball varnum;

ods select position;

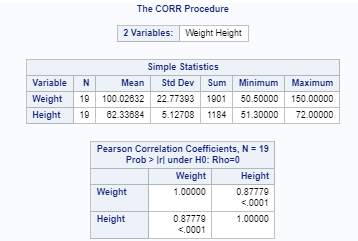
run;



title 'Первые пять Observations Out of 322';  
proc print data=sashelp.Baseball(obs=5);  
run;

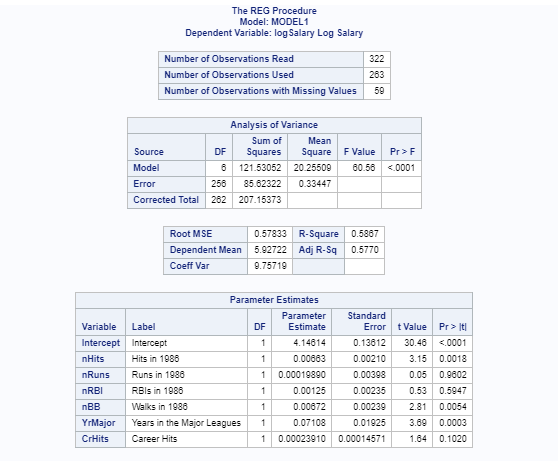


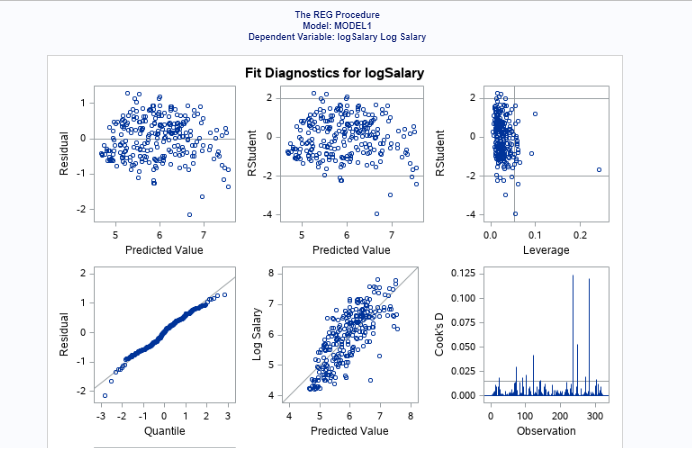
PROC CORR DATA=sashelp.class;  
 VAR weight height;  
RUN;

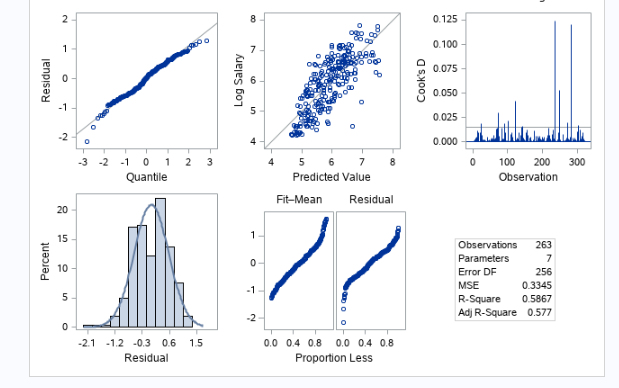


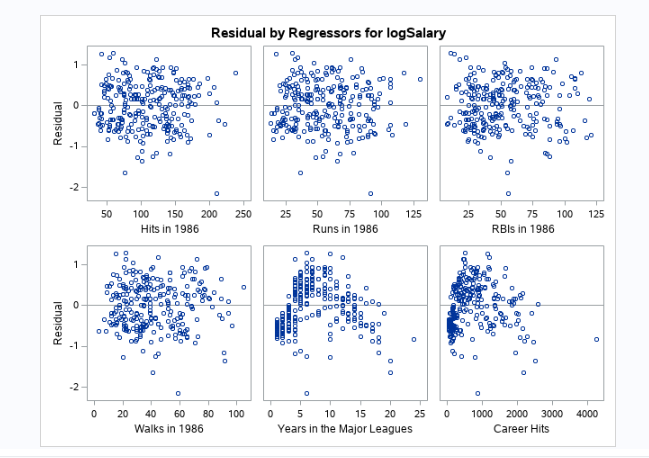
In general is  there a difference between divisions by my criterion?  
  
[2] REGRESSION (further we will build regression and the player’s salary will be our target (dependent) variable)  
[2.1] We will look at the numerical variables in a set (will consider that they the continuous) relating to 1987 (begin on n..., only 8 variables):

ods graphics on;  
  
proc reg data=sashelp.baseball;  
 id name team league;  
 model logSalary = nhits nruns nrbi nbb yrmajor crhits;  
run;



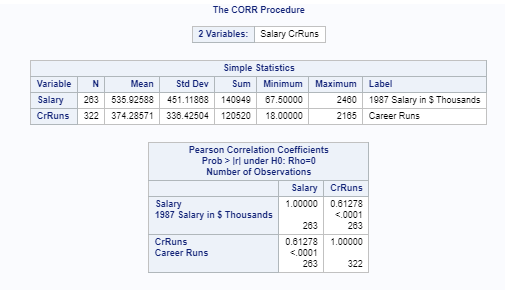




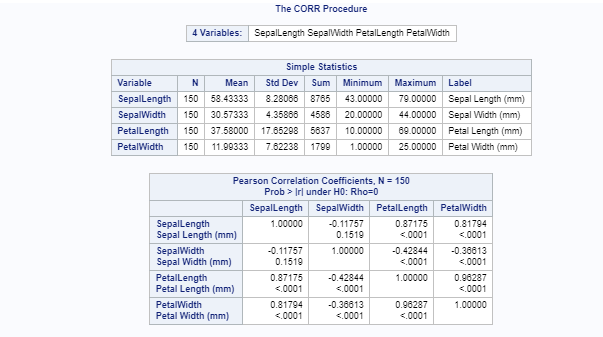


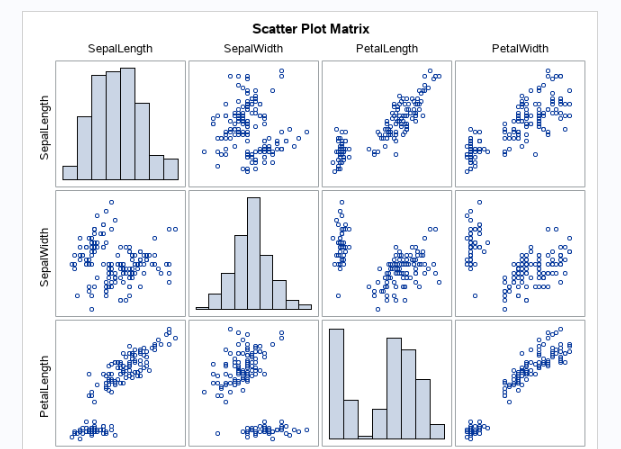
What 3 variables most correlate with target Salary? Specify correlations coefficients for everyone. Specify appropriate p-value?

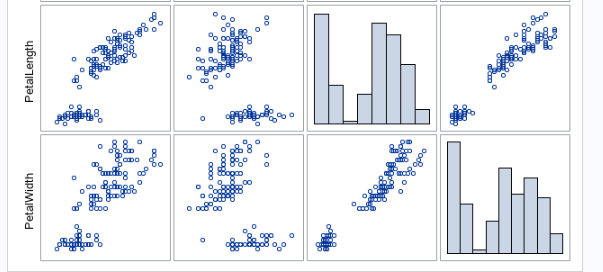
Proc corr data=sashelp.baseball;  
 var Salary CrRuns;  
Run;

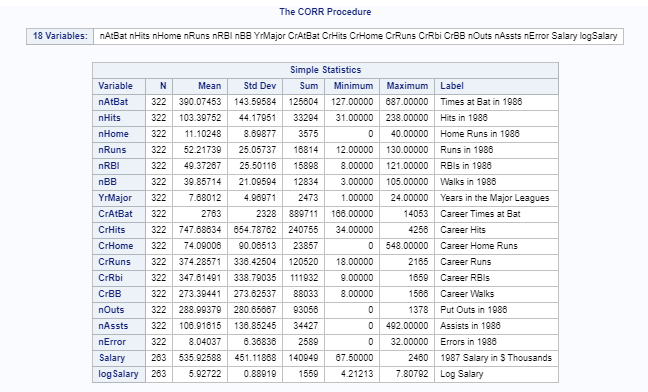


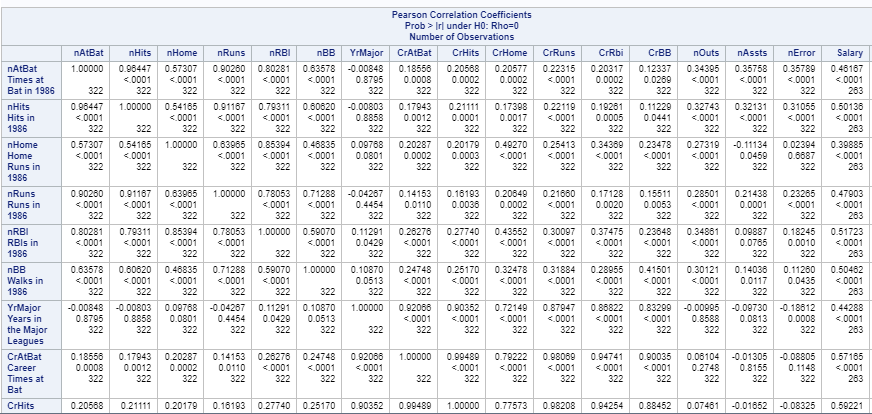
Which 2 variables can be cleaned at once from model?   
proc corr data=sashelp.iris plots=matrix(histogram);  
run;

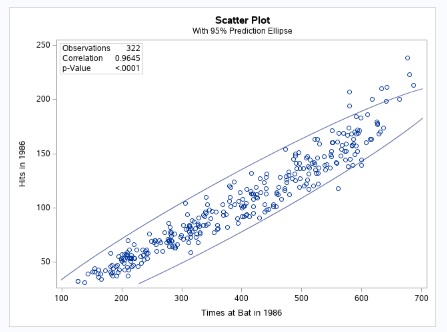


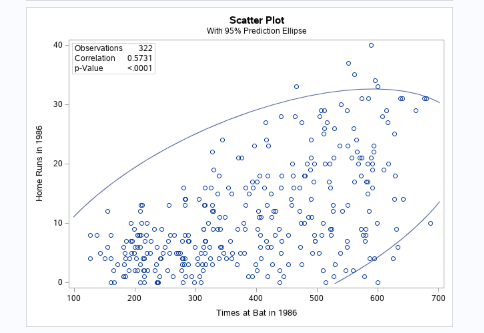


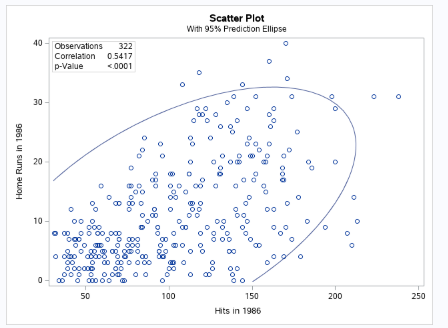
  
May we claim that all calculated values are significant? Why? Whether for any significance value?  
Proc Corr data=sashelp.baseball Plots=scatter;  
run;



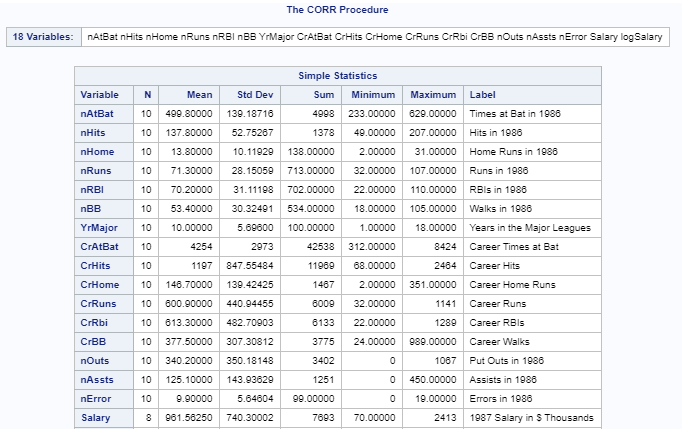


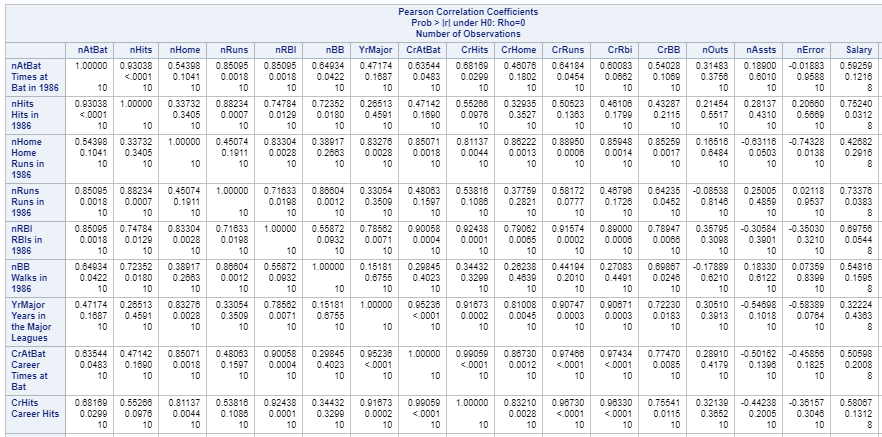


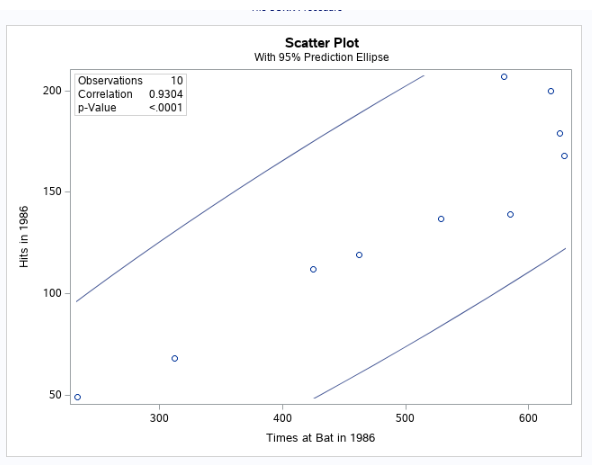


  
 Construct regression using all 6 remained predictors.

Proc Corr data=sashelp.baseball (where=(team='Boston')) plots=all;  
run;







Which predictor is the strongest?

proc contents varnum data=sashelp.baseball;  
 ods select position;  
run;

