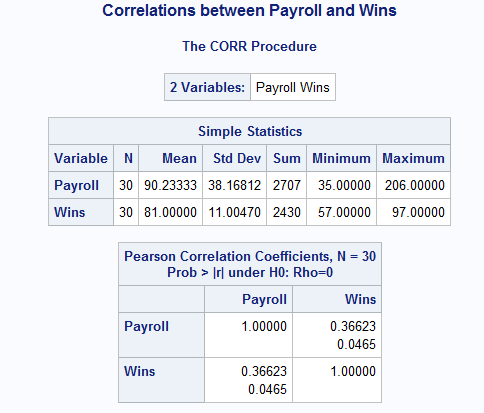
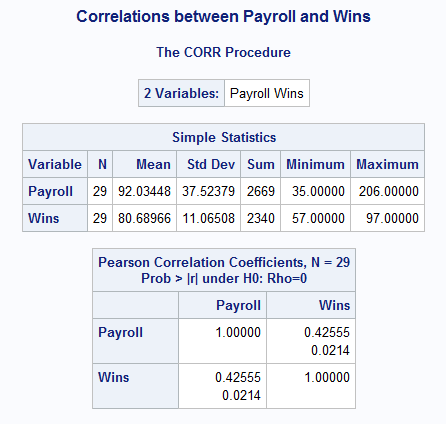


1. Based on the scatter plot, I expect the relationship to be a positive relationship. Meaning, the higher the payroll, the more wins a team can expect. This is because the data trend up and to the right, as payroll increases, wins increase. I would expect a moderate relationship to be established. This is not a completely linear relationship, so the correlation coefficient may not explain the relationship as accurately as possible. I’d guess the correlation coefficient to be around 0.5.



1. The correlation between payroll and wins is 0.37, which is a moderate relationship.
2. By removing the San Diego Padres, our correlation coefficient, r, increases to 0.43:



1. Based on the r-squared statistic of 0.14 on the original data set, payroll only explains roughly 14 % of the change in wins. The league commissioner’s comment has some merit, however, there is a positive relationship associated with team payroll and wins. A better analysis would include years of payroll and wins data, however, the commissioner is correct in his analysis, but it does not PROVE that payroll has no effect on the number of wins. Our analysis based on the correlation coefficient also does not prove causation that payroll hikes will improve wins, it simply states there is a moderate positive relationship between the two variables.
2. The population for these data is the all-encompassing wins and payroll for baseball history. This is a sample of one year from many years of payroll and wins data we could’ve used.

SAS Code:

**data** baseball ;

INPUT Team $ Payroll Wins;

DATALINES;

NYY 206 95

BOS 162 89

CHC 146 75

PHI 142 97

NYM 134 79

DET 123 81

CHW 106 88

LAA 105 80

SF 99 92

MIN 98 94

LAD 95 80

HOU 92 76

SEA 86 61

STL 86 86

ATL 84 91

COL 84 83

BAL 82 66

MIL 81 77

TB 72 96

CIN 71 91

KC 71 67

TOR 62 85

ARZ 61 65

CLE 61 69

WAS 61 69

FA 57 80

TEX 55 90

OAK 52 81

SD 38 90

PIT 35 57

;

**RUN**;

**proc** **print** data = baseball;

**RUN**;

GOPTIONS RESET = ALL;

Title "Baseball Scatter";

**PROC** **GPLOT** data = baseball;

PLOT wins\*payroll;

SYMBOL1 V = Dot C = BLACK I = RL L =**1**;

**RUN**;

**QUIT**;

**PROC** **SORT** data = baseball;

by team;

**RUN**;

**PROC** **CORR** data = baseball;

VAR Payroll Wins;

Title' Correlations between Payroll and Wins';

**RUN**;

**data** baseball2;

INPUT Team $ Payroll Wins;

DATALINES;

NYY 206 95

BOS 162 89

CHC 146 75

PHI 142 97

NYM 134 79

DET 123 81

CHW 106 88

LAA 105 80

SF 99 92

MIN 98 94

LAD 95 80

HOU 92 76

SEA 86 61

STL 86 86

ATL 84 91

COL 84 83

BAL 82 66

MIL 81 77

TB 72 96

CIN 71 91

KC 71 67

TOR 62 85

ARZ 61 65

CLE 61 69

WAS 61 69

FA 57 80

TEX 55 90

OAK 52 81

PIT 35 57

;

**RUN**;

**PROC** **CORR** data = baseball2;

VAR Payroll Wins;

Title' Correlations between Payroll and Wins';

**RUN**;