Tae Coding

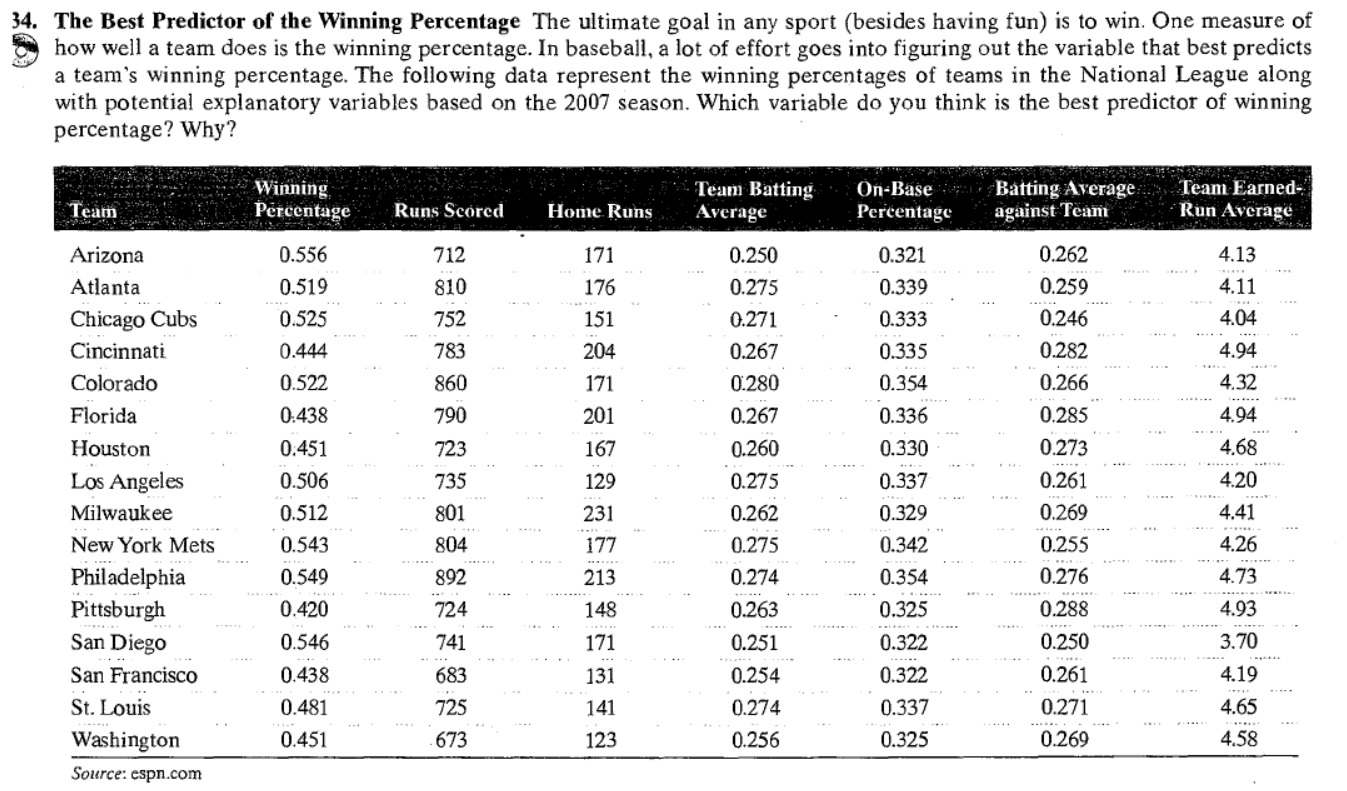
Introduction to Data Science: CS61

Summer 2018

Class Exercise#4.1

Date Given: June 21, 2018 Due Date:

=======================================================================

****

|  |
| --- |
| > #  > # Problem#5  > #  > #########################################  >  > d = read.csv("P5 Book34 Baseball.csv")  > d  X. wP rS hR tbA obP baaT terA  1 1 0.556 712 171 0.250 0.321 0.262 4.13  2 2 0.519 810 176 0.275 0.339 0.259 4.11  3 3 0.525 752 151 0.271 0.333 0.246 4.04  4 4 0.444 783 204 0.267 0.335 0.282 4.94  5 5 0.522 860 171 0.280 0.354 0.266 4.32  6 6 0.438 790 201 0.267 0.336 0.285 4.94  7 7 0.451 723 167 0.260 0.330 0.273 4.68  8 8 0.506 735 129 0.275 0.337 0.261 4.20  9 9 0.512 801 231 0.262 0.329 0.269 4.41  10 10 0.543 804 177 0.275 0.342 0.255 4.26  11 11 0.549 892 213 0.274 0.354 0.276 4.73  12 12 0.420 724 148 0.263 0.325 0.288 4.93  13 13 0.546 741 171 0.251 0.322 0.250 3.70  14 14 0.438 683 131 0.254 0.322 0.261 4.19  15 15 0.481 725 141 0.274 0.337 0.271 4.65  16 16 0.451 673 123 0.256 0.325 0.269 4.58  > cor(d)  X. wP rS hR tbA obP baaT terA  X. 1.0000000 -0.2966325 -0.3404943 -0.3693936 -0.22338385 -0.1976249 0.11708558 0.1225991  wP -0.2966325 1.0000000 0.4590120 0.2674436 0.19118507 0.3276281 -0.65986758 -0.6679483  rS -0.3404943 0.4590120 1.0000000 0.7223473 0.66570447 0.8545529 0.10707850 0.1168089  hR -0.3693936 0.2674436 0.7223473 1.0000000 0.13007580 0.3390246 0.28399732 0.2395204  tbA -0.2233838 0.1911851 0.6657045 0.1300758 1.00000000 0.8816513 0.03708549 0.1781529  obP -0.1976249 0.3276281 0.8545529 0.3390246 0.88165132 1.0000000 0.10303199 0.2040405  baaT 0.1170856 -0.6598676 0.1070785 0.2839973 0.03708549 0.1030320 1.00000000 0.9411114  terA 0.1225991 -0.6679483 0.1168089 0.2395204 0.17815285 0.2040405 0.94111137 1.0000000 |
|  |
| |  | | --- | | > | |

Correlation varies between “-1” and “+1”. A variable which has the highest correlation (positive or negative) with the “Winning Percentage” variable would be the best predictor. A perfect predictor variable will have correlation of either +1 or -1.

Since the correlation of ‘Winning Percentage’ has the highest correlation (-0.667) with the ‘Team Earned Run Average’, it is the best predictor of ‘Winning Percentage’.