STAT-113 Worksheet

**Drive for Show, Putt for Dough?**

Golf is a sport where you hit a ball into a series of holes with as few strokes as possible. The two most important strokes are the drive and the putt. The drive is the first stroke used to hit the ball as far as possible from the tee box towards the hole, while the putt is the final stroke used to roll the ball into the hole on the green.

The PGA2022 data set has statistics with variables for driving (Average Driving Distance, Driving Accuracy Percentage, and Strokes Gained Off the Tee), putting (Average Putts Per Round, One Putt Percentage, and Strokes Gained Putts), and measured success (Scoring Average, Money, and FedexCup Points) with other variables such as Player Name, Country, and Tournament.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | avgDriveDist | drivePct | driveSG | avgPuttsPerRound | onePuttPct | puttsSG | avgScore | Money | Points |
| avgDriveDist | 1.000 | -0.259 | 0.290 | 0.113 | -0.085 | -0.036 | -0.179 | 0.059 | 0.084 |
| drivePct | -0.259 | 1.000 | 0.407 | 0.117 | -0.116 | -0.033 | -0.223 | 0.073 | 0.106 |
| driveSG | 0.290 | 0.407 | 1.000 | 0.192 | -0.172 | -0.140 | -0.334 | 0.207 | 0.249 |
| avgPuttsPerRound | 0.113 | 0.117 | 0.192 | 1.000 | -0.900 | -0.512 | 0.027 | -0.168 | -0.188 |
| onePuttPct | -0.085 | -0.117 | -0.172 | -0.900 | 1.000 | 0.521 | -0.007 | 0.165 | 0.179 |
| puttsSG | -0.036 | -0.033 | -0.139 | -0.512 | 0.521 | 1.000 | -0.358 | 0.298 | 0.319 |
| avgScore | -0.179 | -0.223 | -0.334 | 0.027 | -0.007 | -0.358 | 1.000 | -0.491 | -0.542 |
| Money | 0.059 | 0.073 | 0.207 | -0.168 | 0.165 | 0.298 | -0.491 | 1.000 | 0.950 |
| Points | 0.084 | 0.106 | 0.249 | -0.188 | 0.179 | 0.319 | -0.542 | 0.950 | 1.000 |

1. Which is a better linear predictor of money, average driving distance or average putts per round? Report the correlation.

Average Putts Per Round. R= -0.168

1. Which is a better linear predictor of average score, driving strokes gained or putting strokes gained? Report the correlation.

Strokes gained putt. R= -0.358

1. Which is a better linear predictor of Fedex points, driving percentage or one putt percentage? Report the correlation.

One putt percentage. R= 0.179

1. Look at the correlation matrix above between the Drive Variables and the Success Variables and do the same between the Putt Variables and the Success Variables.
   1. Between which drive variable and success variable is the correlation the strongest? Report the correlation.

Strokes Gained Drive and Average Score. R= -0.333

* 1. Also, between which putt variable and success variable is the correlation the strongest? Report the correlation.

Strokes Gained Putt and Average Score. R= -0.358

1. Write a brief interpretation of the correlation between the drive variable and success variable and the correlation between the putt variable and success variable that you found in #4.

There is a weak negative correlation between Strokes Gained Drive and Average Score, such that as Strokes Gained Drive increases, Average Score decreases. There is a weak negative correlation between Strokes Gained Putts and Average Score, such as Strokes Gained Putts increases, Average Score decreases.

1. Overall, which success variable (Points, avgScore, and Money) is the most predictable and why?

avgScore because it has the best correlated values for drive variables and putt variables

1. In general, what is a better indicator of success, a player’s ability to drive a golf ball or their ability to putt a golf ball? Why do you think this?

Putting because the correlation between the putt variables and the success variables are much greater than the correlation between the drive variables and the success variables.