

STA 4702/5701 Chapter 1 – Data Analysis Problems

LPGA Golf Performance Data

Units: $n = 156$ golfers

Variables: $p = 4$ Characteristics: Drive Distance (X_1), Fairway % (X_2), Green in Regulation % (X_3), Putts/Round (X_4)

Mean		COV_n	Drive	Fairway	Green	Putt
Drive	246.845	Drive	86.233	-25.414	15.602	-1.717
Fairway	67.696	Fairway	-25.414	31.157	4.604	1.120
Green	63.062	Green	15.602	4.604	12.953	0.197
Putt	29.189	Putt	-1.717	1.120	0.197	1.068

1. Compute the bivariate correlations between: Drive (X_1) and Fairway (X_2); and Drive (X_1) and Putt (X_4).

2. Annika Sorrenstam ($j=153$) had:

Drive $x_{153,1} = 245.3$, Fairway $x_{153,2} = 66.2$, Green $x_{153,3} = 62$, and Putt $x_{153,4} = 29.63$.

 - a. Compute her contributions to \bar{x}_1 , $w_{11} = \sum_{j=1}^n (x_{j1} - \bar{x}_1)^2$, $w_{14} = \sum_{j=1}^n (x_{j1} - \bar{x}_1)(x_{j4} - \bar{x}_4)$
 - b. Convert her scores to centered values, then standardized values.
 - c. Compute her statistical distance to the origin (of centered values).

3. Consider only Drive and Fairway for Annika Sorrenstam.
 - a. Rotate her centered scores by $\theta = -21.35^\circ$ to:

$$\tilde{x}_{153,1} = x_{153,1} \cos(\theta) + x_{153,2} \sin(\theta) \quad \text{and} \quad \tilde{x}_{153,2} = -x_{153,1} \sin(\theta) + x_{153,2} \cos(\theta)$$
 - b. The variances of the centered and rotated Drive and Fairway values are given below. Compute her distance from $P_{153} = \left(\tilde{x}_{153,1}, \tilde{x}_{153,2} \right)$ to the origin