

確率と統計（予想問題3）中山クラス

*教科書，ノート，電卓，PC類は使用不可

*答えは回答用紙の所定欄に記入すること

問題 I

Excel等を用いて下記の表を作成し，a.csvとして保存し，read.csv(“a.csv”)によりaaという変数に読み込んだものとする．以下に示す(1)~(8)の命令を実行したときに得られる結果を示せ．また，(9)に関しては文章で述べよ．図を示す場合は概略図でよい．

学生氏名	数学	英語	国語の点数	社会の点数
A	好き	好き	90	80
B	嫌い	嫌い	75	70
C	好き	嫌い	60	65
D	好き	好き	65	60
E	嫌い	嫌い	80	90

- (1) `> aa$数学`
- (2) `> aa[2,]`
- (3) `> mean(aa[,4])`
- (4) `> table(aa[,3])`
- (5) `> table(aa$数学, aa$英語)`
- (6) `> var(aa[,5])`
- (7) `> cov(aa$国語の点数, aa$社会の点数)`
- (8) `> plot(aa[,4], aa[,5])`
- (9) (6)の結果に基づき，国語の点数と社会の点数の間にはどのような相関性があるかのべよ．

問題 II

統計量に関して以下の問に答えよ．

- (1) 相関係数とファイ係数の違いを述べよ．
- (2) 標本分散と不偏分散の違いを述べよ．
- (3) 共分散と相関係数の違いを述べよ．

問題 III

共分散及び相関係数に関して以下の問に答えよ．データを $X = [x_1, x_2, \dots, x_n]$, $Y = [y_1, y_2, \dots, y_n]$ とし，各々の平均を μ_x, μ_y とする．

- (1) 共分散の計算において平均を引く理由を述べよ． $(x_i - \mu_x), (y_i - \mu_y)$
- (2) 共分散の計算において積を用いる理由を述べよ． $(x_i - \mu_x)(y_i - \mu_y)$
- (3) 共分散において平均を計算する理由を述べよ． $\frac{1}{n} \sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)$
- (4) 相関係数は共分散を標準偏差で割り算して求められるが，その理由を述べよ．