確率と統計　小テスト（中山・金曜クラス）答案用紙

2012.5.25(金)

配点：正解（○）：1 点／設問，解答が正解に近い場合（△）：0.5 点／設問
合計点：21 点（答案用紙に記載）→　成績評価の際に 40 点満点に換算

問題 I

(1) 成績の度数分布表を作成せよ．

A B C D

3 4 2 1

(2) 成績の最頻値を求めよ．

B

(3) 試験点数の平均を求めよ．

70

(4) 試験点数の標本分散を求めよ．

120

(5) 試験点数の標準偏差（標本分散による）を求めよ．

11.0

(6) 学生 ID=3 の試験点数の z 得点を求めよ．

0.9

(7) 学生 ID=3 の試験点数の偏差値を求めよ．

59.1

問題 II

＜命令文と実行結果の双方が正しい場合：○，片方のみ正しい場合：△＞

(1) ヘッダーが英語である列を表示する．

```
> aa$英語
[1] 好き 嫌い 嫌い 好き 嫌い
Levels: 嫌い 好き
> aa[,3]
[1] 好き 嫌い 嫌い 好き 嫌い
```

```
Levels: 嫌い 好き
```
＜上記の２通りの方法のいずれかでよい．他の設問でも同じ＞

(2) 第３行目（ヘッダーを除く）を表示する．

```
> aa[3,]
  学生氏名 数学 英語 国語の点数 社会の点数
3        C 嫌い 嫌い         60         70
```

(3) 社会の点数の平均を計算する．

```
> mean(aa$社会の点数)
[1] 60
> mean(aa[,5])
[1] 60
```

(4) 数学の好き／嫌いの度数分布表を求める．

```
> table(aa$数学)
嫌い 好き
   2    3
> table(aa[,2])
嫌い 好き
   2    3
```

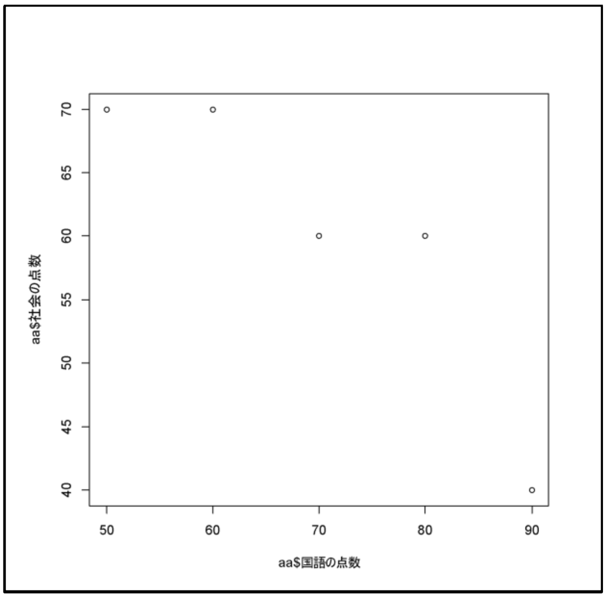
(5) 数学と英語の好き／嫌いのクロス集計表を求める．

```
> table(aa$数学, aa$英語)
       嫌い 好き
  嫌い    2    0
  好き    1    2
> table(aa[,2], aa[,3])
       嫌い 好き
  嫌い    2    0
  好き    1    2
```

(6) 国語の点数と社会の点数の散布図を求める．

右の枠内に散布図を示せ．

```
> plot(aa$国語の点数, aa$社会の点数)
> plot(aa[,4], aa[,5])
```



(7) (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があるか答えよ．

散布図が右下がりであるので，負の相関がある．

問題Ⅲ

(1) 量的変数と質的変数の違いを述べよ．

量的変数はデータが数値で与えられるもの．質的変数とは「好き」「嫌い」など数値では表現できないもの．

(2) 散布図とクロス集計表の違いを述べよ．

散布図は２つの量的変数の関係をグラフであらわしたもの．横軸と縦軸に量的変数の値をとる．グラフが右上がりの場合が正の相関，右下がりの場合が負の相関を表す．

クロス集計表は２つの質的変数の関係を表で表したもの．縦方向と横方向に質的変数の場合分け（例：好き，嫌い）をとり，それに対応する度数を表に記入する．斜め方向に大きな度数がある場合は負または正の連関がある．

(3) ヒストグラムと度数分布表の違いを述べよ．

ヒストグラムは量的変数xの分布を棒グラフで表したもの．xが$a \leq x \leq b$の範囲にあるデータの数を棒グラフとして表す．度数分布表は質的変数の分布を表で表したもの．例えば，ある科目の成績で A をとった人数，B をとった人数を表の形式で表現する．

問題Ⅳ

＜以下のような内容が一部でも含まれていれば正解とする＞

(1) 共分散の計算において平均を引く理由を述べよ．$(x_i - \mu_x)$，$(y_i - \mu_y)$

共分散は２つの変数の相関を表す．x_i, y_iに共通に含まれる成分は相関には無関係な成分であるので，これを除く必要がある．平均を引くことにより，x_i，y_iの関係を強調できる．

(2) 共分散の計算において積を用いる理由を述べよ．$(x_i - \mu_x)(y_i - \mu_y)$

$(x_i - \mu_x)$，$(y_i - \mu_y)$は正の相関がある場合は大半が（正，正）または（負，負）の値をとる．一方，負の相関がある場合は大半が（正，負）または（負，正）の値となる．これらは，積をとることにより，前者の大半が正の値，後者の大半が負の値となり相関をより強調できる．

(3) 共分散において平均を計算する理由を述べよ．　$\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)$

相関を一つの数値で表現するために，データを集約する必要がある．データの集約方法として(2)で述べた積の平均を計算することにより，正または負の相関に対しては正または負の値となり，相関が低い場合は絶対値が小さな値となる．

(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．

共分散はデータの大きさによって，その大きさが変動する．そのために，共分散の大きさが相関の強さに対応しない．共分散をデータの大きさである標準偏差で割る（正規化する）ことにより，データに大きさに関係なく，相関の強さのみを表現できる．相関係数はデータの大きさに関係なく-1～1の値を取り，その大きさは相関の強さのみを表す．