

確率と統計 小テスト (中山・火曜クラス) 解答例

2012.5.22(火)

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：25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 (実行結果におけるデータの並び方や順番は問わないものとする)

(1) `> aa$数学`

[1] 好き 好き 嫌い 好き 嫌い ([1]はなくてもよい)

Levels: 嫌い 好き (この部分はなくてもよい)

(2) `> aa[2,]`

学生氏名 数学 英語 国語の点数 社会の点数

2 B 好き 嫌い 70 65

(3) `> mean(aa[,4])`

[1] 70

(4) `> table(aa[,3])`

嫌い 好き

3 2

(5) `> table(aa$数学, aa$英語)`

嫌い 好き

嫌い 2 0

好き 1 2

(6) (5)の結果から，

(a) 数学と英語の好き嫌いの間にはどのような関係があるかを述べよ。

数学が嫌いな2人は同時に英語が嫌いであり，また，数学が好きな3人の内，2人は同時に英語も好きである．以上より，数学と英語の好き嫌いには連関があると言える．

(b) ファイ係数はどのような値になるか予想せよ．

数学が嫌い(=1)，好き(=0)，英語が嫌い(=1)，好き(=0)のように数値化して相関係数を計算することによりファイ係数が計算される．(5)の結果より，「数学が嫌い(=1)と英語が嫌い(=1)，及び，「数学が好き(=0)と英語が好き(=0)」の連関が強いので，ファイ係数も正の比較的大きな値になると考えられる．実際に計算すると0.67となる．従って，ファイ係数として「0.3～1」の範囲を正解とする．尚，実際に計算する必要はなく，予想される数値を記入すればよい．

(7) `> var(aa[,5])`

[1] 250

(8) `> sd(aa[,4])` 小数点以下 2 桁目を四捨五入する. $\sqrt{10} = 3.16$

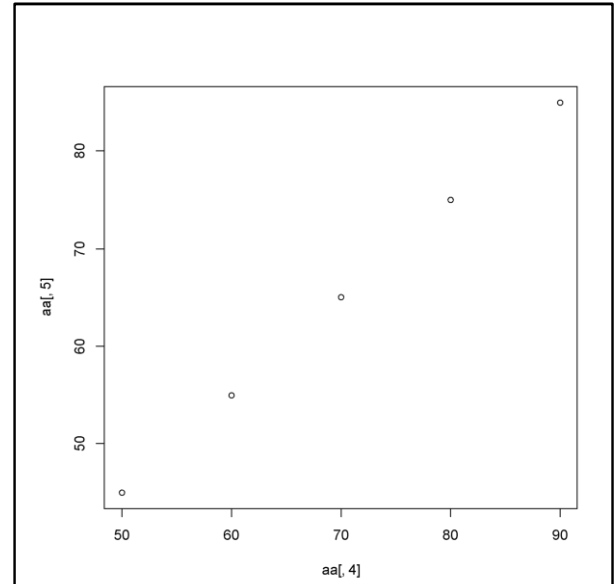
[1] 15.81139

解答は 15.8 でよい.

(9) `> cov(aa$国語の点数, aa$社会の点数)`

[1] 250

(10) 右の四角の枠内にプロット図を示せ.



(11) (10)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.

プロットは右上がりの直線上にあり, 正の強い相関がある.

(12) (a) 数学の好き嫌いとは国語の点数の間にはどのような関係があるか答えよ.

数学の好きな人は国語の点数が高く, 嫌いな人は低い. 従って, 数学が好き=1, 数学が嫌い=0 とすると, 数学の好き嫌いとは国語の点数の間には正の相関があると言える.

(b) 英語の好き嫌いとは社会の点数の間にはどのような関係があるか答えよ.

英語が好きな人は社会の点数が高く, 嫌いな人は点数が低い. (a)と同様な計算をすると, 英語の好き嫌いとは社会の点数の間には正の相関があると言える.

問題 II

(1) 相関係数とファイ係数の違いを述べよ.

相関係数は量的変数の相関を表す. $-1 \sim 1$ の値を取り, -1 に近いとき, 負の強い相関, 1 に近いとき正の強い相関がある. また, 絶対値が小さいときには相関は弱い. 一方, ファイ係数は質的変数の連関を表す. 一方を 0 , 他方を 1 とおいて, 相関係数と同じ方法で計算されたため, 相関係数と同じ性質を持つ.

(2) 標本分散と不偏分散の違いを述べよ.

全てのデータが分かっているとき, その全てのデータを用いて計算されるのが標本分散であり, 全てのデータが分かっていないときに, 全体のデータに対する分散を推定するのが不偏分散である. 計算式においては, 標本分散の分母はデータの個数であり, 不偏分散の分母はデータの個数-1 である.

(3) 共分散と相関係数の違いを述べよ.

共分散は2つの変数の相関を表すが、データの大きさにより変化する（影響を受ける）ため、その大きさが相関の強さを直接表すことはできない。これに対して、相関係数は共分散を標準偏差（データの大きさ）で割っている（正規化している）ため、データの大きさによって変化せず（影響を受けず）、その大きさが相関の強さを表す。相関係数は-1~1の値をとる。

問題Ⅲ

(1) X と Y の平均 → X_m, Y_m とする。

$$X_m = \frac{1}{n} \sum_{i=1}^n x_i \quad Y_m = \frac{1}{n} \sum_{i=1}^n y_i$$

(2) Y の不偏分散、但し、 Y_m を用いて表せ。

$$\frac{1}{n-1} \sum_{i=1}^n (y_i - Y_m)^2$$

$$\frac{1}{n-1} \sum_{i=1}^n y_i^2 - \frac{n}{n-1} Y_m^2$$

(3) X と Y の標本分散、但し、 X_m, Y_m を用いて表せ。 X_v, Y_v とする。

$$X_v = \frac{1}{n} \sum_{i=1}^n (x_i - X_m)^2 \quad Y_v = \frac{1}{n} \sum_{i=1}^n (y_i - Y_m)^2$$

$$X_v = \frac{1}{n} \sum_{i=1}^n x_i^2 - X_m^2 \quad Y_v = \frac{1}{n} \sum_{i=1}^n y_i^2 - Y_m^2$$

(4) X と Y の標準偏差（標本分散による）、但し、 X_v, Y_v を用いて表せ。 → X_{sd}, Y_{sd} とする。

$$X_{sd} = \sqrt{X_v} \quad Y_{sd} = \sqrt{Y_v}$$

(5) X と Y の共分散、但し、 X_m, Y_m を用いて表せ。 → XY_{cov} とする。

$$XY_{cov} = \frac{1}{n} \sum_{i=1}^n (x_i - X_m)(y_i - Y_m)$$

$$XY_{cov} = \frac{1}{n} \sum_{i=1}^n x_i y_i - X_m Y_m$$

(6) X と Y の相関係数、但し、 X_{sd}, Y_{sd}, XY_{cov} を用いて表せ。

$$\frac{XY_{cov}}{X_{sd} Y_{sd}}$$

(7) x_1 の z 得点、但し、 X_m, X_{sd} を用いて表せ。 → x_{1z} とする。

$$x_{1z} = \frac{x_1 - X_m}{X_{sd}}$$

(8) x_1 の偏差値, 但し, x_{1z} を用いて表せ.

$$10x_{1z} + 50$$