

1. 概要

n 個のデータ対の集合に対して線形回帰パラメータ β_0, β_1 , 相関係数 $r_{x,y}, r^2$ を計算する。

与えられた見積もり値 x_k に対して, $y_k = \beta_0 + \beta_1 x_k$ を満たす予測値 y_k を計算する

2. 詳細

回帰パラメーター β_1, β_0 は x, y の平均 x_{avg}, y_{avg} を用いて以下の式で求められる。

$$\beta_1 = \frac{(\sum_{i=1}^n x_i y_i) - (n x_{avg} y_{avg})}{(\sum_{i=1}^n x_i^2) - (n x_{avg}^2)}$$

$$\beta_0 = y_{avg} - \beta_1 x_{avg}$$

相関係数 $r_{x,y}, r^2$ は以下の式で求められる

$$r_{x,y} = \frac{n(\sum_{i=1}^n x_i y_i) - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{\sqrt{[n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2][n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2]}}$$

$$r^2 = r_{x,y}^2$$

しかし、相関係数 $r_{x,y}$ は x, y の標準偏差 σ_x, σ_y を用いての以下の式に変換できるため、それを利用して以下で求める。

$$r_{x,y} = \beta_1 \frac{\sigma_x}{\sigma_y}$$

ここで n は与えられるデータ対の数、 x_i は i 個目の実数の値を表す。

また、与えられる n 個のデータ対は、双方向リンクリストを用いて操作する。

3. 入力

- データの入力：csv ファイル入力
- 入力ファイル：ファイル例の図 1 のように、カンマ区切りの表形式で表現し、同じ列に同じ属性のデータを格納する。

なお, 与えるデータの各列は以下のデータを表す.

- 1 列: 代替品の規模見積もり
- 2 列: 追加修正の計画規模
- 3 列: 追加修正の実績規模
- 4 列: 実績開発時間

```
test > 課題3 > data_pair.csv
1 130,163,186,15.0
2 650,765,699,69.9
3 99,141,132,6.5
4 150,166,272,22.4
5 128,137,291,28.4
6 302,355,331,65.9
7 95,136,199,19.4
8 945,1206,1890,198.7
9 368,433,788,38.8
10 961,1130,1601,138.2
```

図 1: データ対入力ファイル例

- 実行時の入力: コマンドラインに以下の形式で入力 `java プログラム名 実数値入力ファイル名 x として用いる列の番号 y として用いる列の番号 x_k`
- 実行時入力例: `java Program3 data_pair.csv 2 3 386`

4. 出力

- 出力方法: コマンドライン出力
- 出力する値: 回帰パラメータ β_0, β_1 、相関係数 $r_{x,y}, r^2$ 、予測値 y_k
- 精度: 必要に応じて少数第 7 位を四捨五入して表示する。
- 出力例: 図 2 のように出力する値をそれぞれ `b_1`, `b_2`, `r_xy`, `r2`, `y_k` として改行して表示する。

```
b_0 = -22.55
b_1 = 1.7279
r_xy = 0.9545
r2 = 0.9111
y_k = 644.429
```

図 2: 出力例

5. テスト

図 1 のデータをもちいて、 x, y を以下の 4 つ組み合わせでテストを行い、それぞれの期待値を図 3, 図 4, 図 5, 図 6 に示す。

ただし、代理品の規模見積もりは $x_k=386$ として与えられている。

- テスト 1: x : 代理品の規模見積もり、 y : 追加修正の実績規模
- テスト 2: x : 代理品の規模見積もり、 y : 実績開発時間
- テスト 3: x : 追加修正の計画規模、 y : 追加修正の実績規模
- テスト 4: x : 代理品の規模見積もり、 y : 追加修正の実績規模

```
b_0 = -22.55
b_1 = 1.7279
r_xy = 0.9545
r2 = 0.9111
y_k = 644.429
```

図 3: 期待値 1

```
b_0 = -4.039
b_1 = 0.1681
r_xy = 0.9333
r2 = 0.8711
y_k = 60.858
```

図 4: 期待値 2

```
b_0 = -23.92  
b_1 = 1.43097  
r_xy = 0.9631  
r2 = 0.9276  
y_k = 528.4294
```

图 5: 期待值 3

```
b_0 = -4.604  
b_1 = 0.140164  
r_xy = 0.948  
r2 = 0.8988  
y_k = 49.4994
```

图 6: 期待值 4