数理統計学　前回の解答

次のデータの散布図を描き、相関係数を求めよ。

番号	1	2	3	4	5	6	7	8
x	3	4	5	6	7	8	9	10
y	13.7	22.2	19.5	27.3	29.5	35.1	47.5	55.6

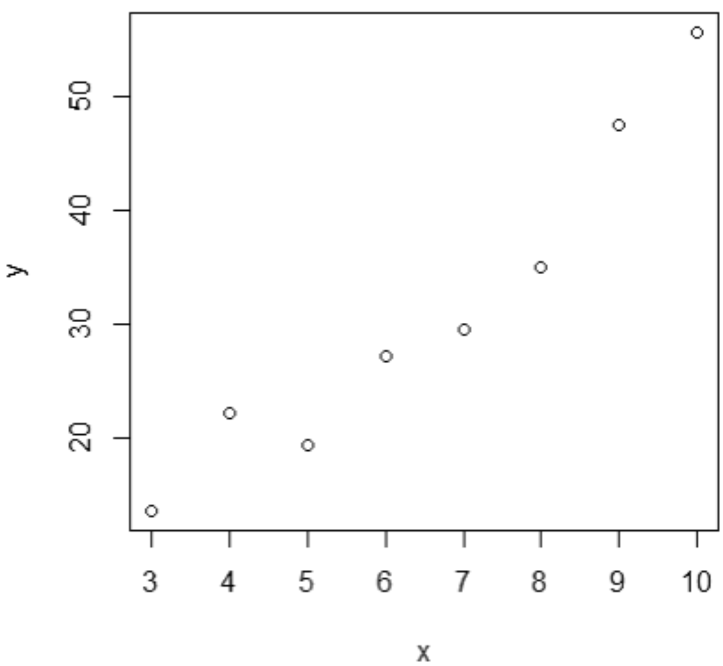
《解答例》R のプログラムを書くとデータと相関係数は

```
> x <- c(3,4,5,6,7,8,9,10)
> y <- c(13.7,22.2,19.5,27.3,29.5,35.1,47.5,55.6)
> rbind(x,y)
  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
x  3.0  4.0  5.0  6.0  7.0  8.0  9.0 10.0
y 13.7 22.2 19.5 27.3 29.5 35.1 47.5 55.6
> cor(x,y)
[1] 0.9603522
```

となり、相関係数は $r = 0.960$ であることがわかる。

また、散布図は次のようになる。

```
> plot(x,y)
```



数理統計学　回帰直線（回帰分析）

・回帰分析について

回帰分析はデータの予測に用いられる手法で、いくつかの説明変数 $x_1, x_2, \cdots, x_p$ から目的変数 y を求める（予測する）関係式 $y = f(x_1, \cdots, x_p)$ を推定する。簡単な例として、おもりの重さとばねばかりの伸びは正比例の関係があることが知られている。つまりおもりの重さを x g、ばねばかりの伸びを y cm としたとき、

$$y = \alpha + \beta x$$

の関係がある（この直線を回帰直線と呼ぶ)。ばねばかりに関する情報が何も無い場合、何回かの実験によって得られた (x_i, y_i) の組で与えられる標本から α, β の値を推定する必要がある。測定にはある程度の誤差が考えられるので、すべての標本が $y = \alpha + \beta x$ 上に現れるわけではない。そこで次のような誤差 ε_i を考慮した関係式

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

を考えて、誤差を最小にする α, β の値を求める。誤差の単純和は正の値と負の値が打ち消しあってしまうため、実際には誤差の 2 乗和を最小にする α, β の値を計算する。詳しい説明は省くが、関数の最大値を求める問題なので、関数を α, β で偏微分し、0 となる α, β を求めればよい。実際に n 個の標本が与えられたとき、$\sum \varepsilon_i^2$ を最小にする α, β の推定値 $\hat{\alpha}, \hat{\beta}$ は次のようになる。

$$\begin{aligned}
\hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x} \\
\hat{\beta} &= \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2} \\
&= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x^2}
\end{aligned}$$

つまり平均および分散と x, y の積和があれば求めることが可能である。ちなみに相関係数はこの回帰直線にどれだけ近いかを表す統計量と考えることも可能である。

・R を使った回帰直線の求め方

もちろん R にも関数が用意されており、次のようにして求めることができる。

例：年齢と血圧

データの入力

```
> x <- c(37,45,48,55,58,67,75)
> y <- c(112,132,120,154,136,162,164)
```

データのプロット（散布図を書く）

```
> plot(x,y)
```

回帰直線の計算（実際には `result` に回帰直線の結果を代入）

```
> result <- lm(y~x)
```

回帰直線を書く（散布図に上書きされるので `par(new=T)` は必要なし）

```
> abline(result)
```

結果の表示（回帰直線 $y = 1.419x + 61.949$ のみが分かる）

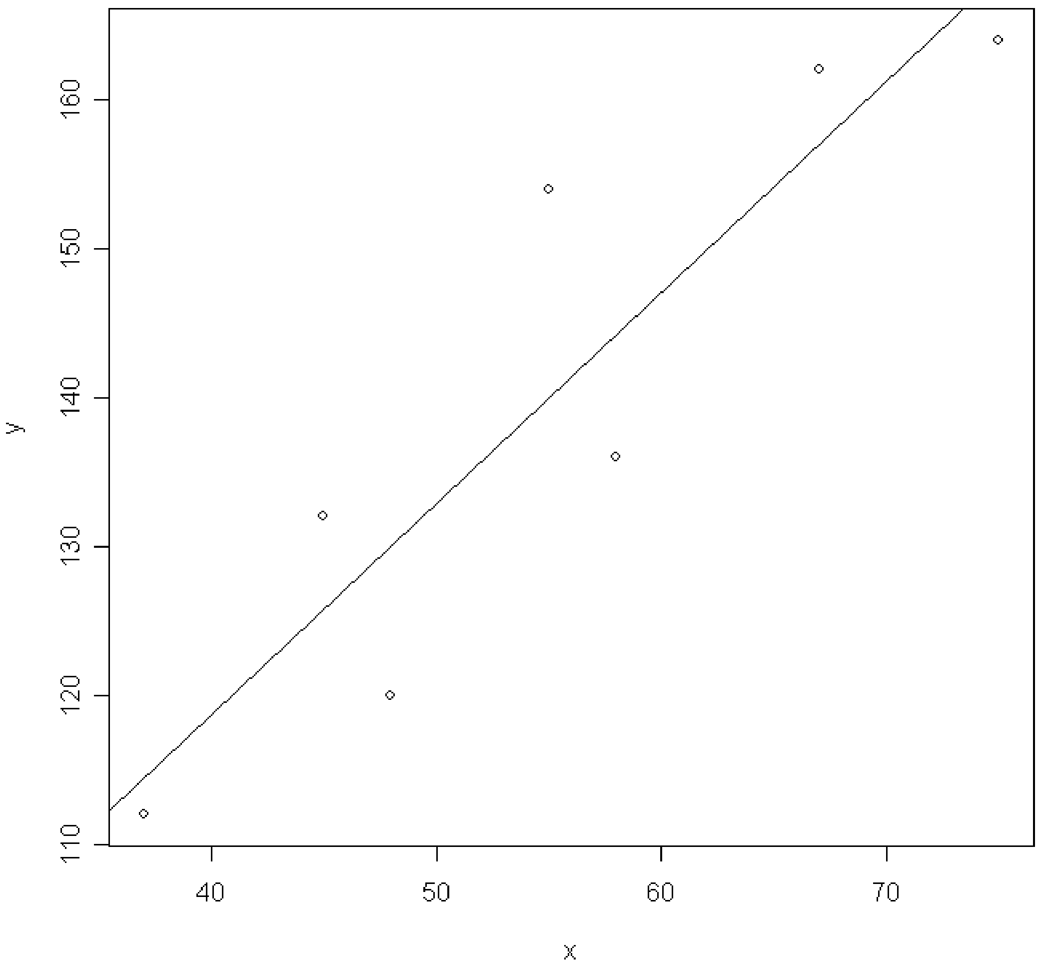
```
> result
Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
    61.949        1.419
```

詳細の表示（いろいろな結果が表示される）

```
> summary(result)

Call:
lm(formula = y ~ x)

Residuals:
      1       2       3       4       5       6       7
 -2.456   6.191 -10.066  14.000  -8.257   4.971  -4.382

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  61.9493    16.7496   3.699  0.01402 *
x             1.4191     0.2974   4.771  0.00501 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.527 on 5 degrees of freedom
Multiple R-Squared: 0.8199,     Adjusted R-squared: 0.7839
F-statistic: 22.77 on 1 and 5 DF,  p-value: 0.005009
```

相関係数の計算

```
> cor(x,y)
[1] 0.9054979
```



2019年度 数理統計学 N科　小テスト解答用紙　2020. 1.15

・相関係数

次のデータの相関係数 r と回帰直線 $y = a + bx$ を求め、散布図に回帰直線を描け。

番号	1	2	3	4	5	6	7	8	9
x	11	13	15	17	19	21	23	25	27
y	99.3	83.5	82.1	72.5	71.6	53.7	41.9	34.3	26.3

2019年度神奈川工科大学 数理統計学 演習問題	学科	学年	組	学　籍　番　号	氏　　　名	

提出先：K3-3309号室前　19番のボックス　提出期限：　1月１６日（木）１７時頃まで