統計学入門　宿題 #2.　2005年夏学期　解答例

1(a) [Set A]　$\bar{x} = \sum_{i=1}^{11} x_i / 11 = 9$　　$\bar{y} = \sum_{i=1}^{11} y_i / 11 = 7.50$

$S_{xy} = \frac{1}{11-1} \sum_{i=1}^{11} (x_i - \bar{x})(y_i - \bar{y}) \approx 5.50$

$S_x = \sqrt{\frac{1}{11-1} \sum_{i=1}^{11} (x_i - \bar{x})^2} \approx 3.32$, $S_y = \sqrt{\frac{1}{11-1} \sum_{i=1}^{11} (y_i - \bar{y})^2} \approx 2.03$

相関係数　$r_{xy} = \frac{S_{xy}}{S_x S_y} \approx 0.82$

回帰直線の傾き　$b = \frac{S_{xy}}{S_x^2} \approx 0.50$

〃　切片　$a = \bar{y} - b\bar{x} = 3.00$　　よって　$y = 3.00 + 0.50x$.

$x = 15$のときのyの予測値　$\hat{y} = \hat{a} + \hat{b} \cdot 15 = 10.50$

[Set B]　Set Aと同様にして,

相関係数　$r_{xy} = \frac{S_{xy}}{S_x S_y} \approx 0.82$.

回帰直線の傾き　$b = \frac{S_{xy}}{S_x^2} \approx 0.50$

〃　切片　$a = \bar{y} - b\bar{x} = 3.00$　　$y = 3.00 + 0.50x$

$x = 15$のときのyの予測値　$\hat{y} = \hat{a} + \hat{b} \cdot 15 \coloneqq 10.50$

[Set C]　相関係数　$r_{xy} = \frac{S_{xy}}{S_x S_y} \approx 0.82$.

回帰直線の傾き　$b = \frac{S_{xy}}{S_x^2} \approx 0.50$

〃　切片　$a = \bar{y} - b\bar{x} \approx 3.00$　　$y = 3.00 + 0.50x$.

$x = 15$のときのyの予測値　$\hat{y} = \hat{a} + \hat{b} \cdot 15 = 10.50$

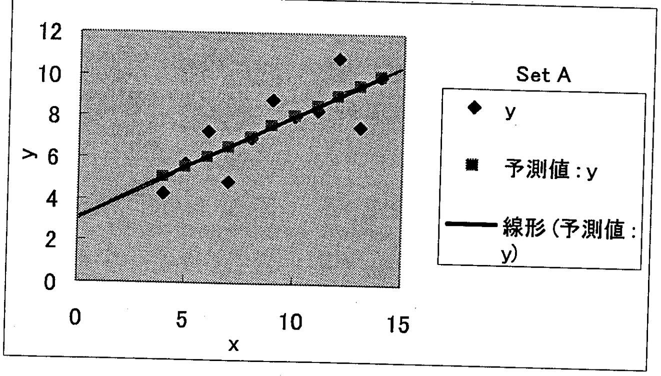
[Set D]　相関係数　$r_{xy} = \frac{S_{xy}}{S_x S_y} \approx 0.82$

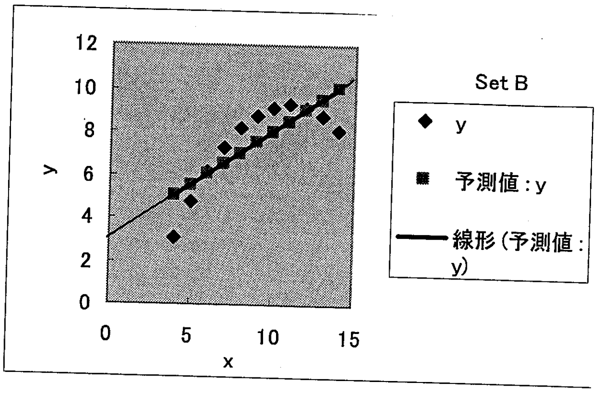
回帰直線の傾き　$b = \frac{S_{xy}}{S_x^2} \approx 0.50$

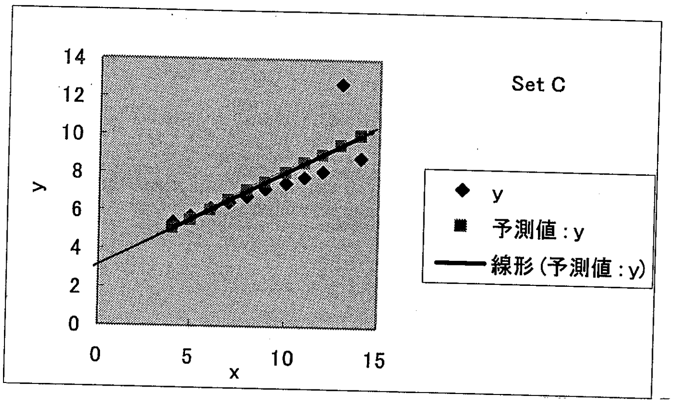
〃　切片　$a = \bar{y} - b\bar{x} \approx 3.00$　　$y = 3.00 + 0.50x$.

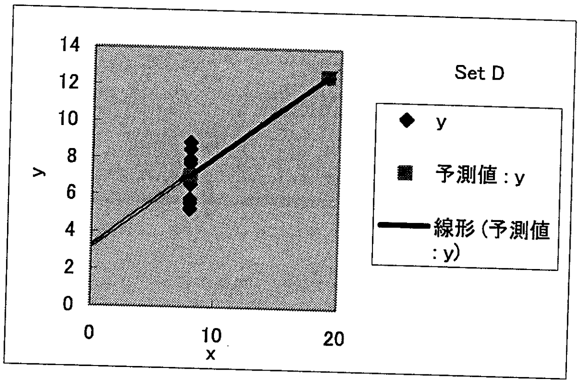
$x = 15$のときのyの予測値　$\hat{y} = \hat{a} + \hat{b} \cdot 15 = 10.50$

1.(b)









1.(c) Set A : 説明している

Set B : 説明していない (非線形回帰を行うべきである。)

Set C : 説明していない (異常値をとりのぞいて回帰を行うべきである)

Set D : 説明していない (回帰分析は無理である。)

1.(d) 単回帰のばあい、決定係数 R^2 は相関係数 r_{xy} の2乗に等しい。

よって

Set A : $R^2 = r_{xy}^2 \approx 0.67$

Set B : $R^2 = r_{xy}^2 \approx 0.67$

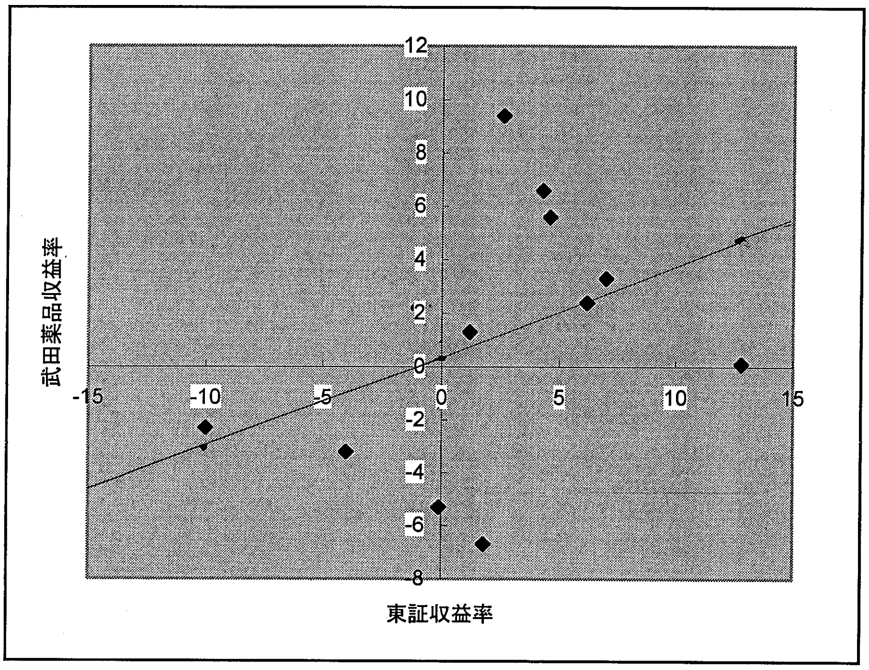
Set C : $R^2 = r_{xy}^2 \approx 0.67$

Set D : $R^2 = r_{xy}^2 \approx 0.67$

＊できるだけ、1ケタより大きくなるよう、四捨五入して下さい。1ケタではあらすぎます。

ex. 相関係数 0.8 でなく 0.82.

ex. 回帰直線 $3 + 0.5x$ でなく、$3.0 + 0.50x$.



2.

東証収益率を x,
武田薬品収益率を y とおく。
x_i $(i=1,2,\cdots 12)$ と
y_i $(i=1,2,\cdots 12)$ の
散布図が左の図である。
このとき、単回帰分析をおこなう
回帰式は,
$y = a + bx + \text{error}$
である.

$\bar{x} = \frac{1}{12}\sum_{i=1}^{12} x_i \approx 2.12$

$\bar{y} = \frac{1}{12}\sum_{i=1}^{12} y_i \approx 1.12$

相関係数 $r_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{\frac{1}{11}\cdot 113.1}{\sqrt{\frac{1}{11}\cdot 365.6}\cdot\sqrt{\frac{1}{11}260.0}} \approx 0.367$

回帰式の傾き $b = \frac{S_{xy}}{S_x^2} = \frac{\frac{1}{11}\cdot 113.1}{\left(\sqrt{\frac{1}{11}\cdot 365.6}\right)^2} \approx 0.309$

回帰式の切片 $a = \bar{y} - b\bar{x} \approx 0.462$

よって回帰式は. $y = 0.462 + 0.309x$ となる。

決定係数 R^2 は, 単回帰のばあい 相関係数の2乗であるから.

$$R^2 = r_{xy}^2 = \left(\frac{S_{xy}}{S_x S_y}\right)^2 \approx 0.135$$

*. 何を x とし, 何を y にするのか, 書いていない答案がめだちました。
きちんと 書きましょう

回帰のあてはまりは. あまりよくない。

統計学入門　宿題＃2　2005年夏学期、解答例（補足）

回帰直線のあてはまりについて

。残差2乗和が0に近ければあてはまりが良いと判断している解答が目立ちました。すなわち、

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

という単回帰分析をおこない、最小2乗法によりえられたαとβの値をそれぞれa, bとおく。このときの残差は、

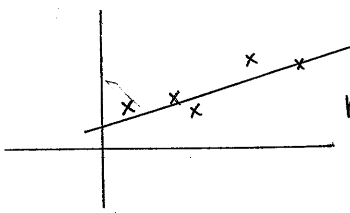
$$e_i = y_i - \hat{y}_i = y_i - (a + b x_i) \qquad i = (1, \cdots m)$$

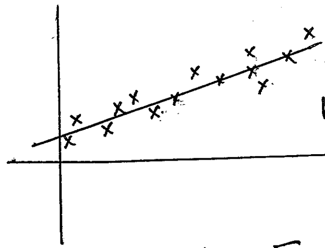
となります。最小2乗法において残差の和 $\sum_{i=1}^{m} e_i$ は、0です。

残差2乗和は、

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

ですが、この値はnが大きくなるにつれて、大きくなります。つまり、$\sum_{i=1}^{n} e_i^2$が0に近いかどうかということは、nに左右され、しかも、0と近いかどうかは、





個人の基準で判断するしかありません。（例えば、$\sum e_i^2 = 10$が0と近いかどうか ???）　決定係数は、

$$R^2 = \frac{回帰変動}{全変動} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

$$= 1 - \frac{残差変動}{全変動} = 1 - \frac{\sum e_i^2}{\sum(y_i - \bar{y})^2}$$

ですから、残差2乗和を用いるよりもR^2を用いた方が良いわけです。

但し、問題1のように、グラフ等から視覚的に残差e_iのばらつきを確認することも大切です。