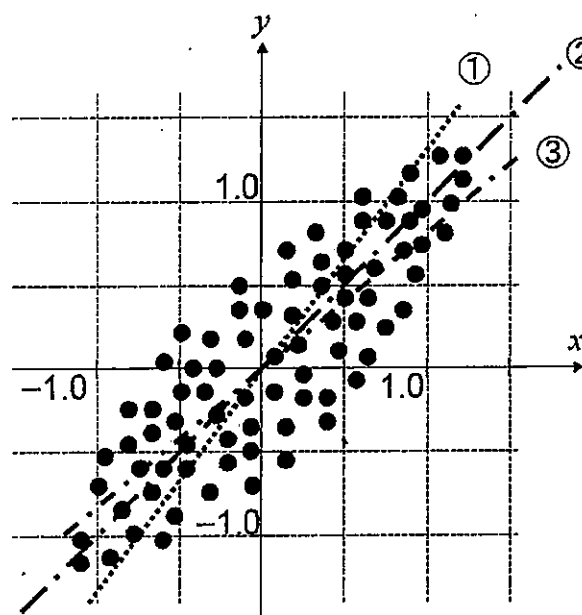


## 情報生物学 1/1

- [1] 雌の親ネズミ  $n$  個体の体長を  $(X_1, X_2, \dots, X_n)$ 、それらの雌の子ネズミ  $n$  個体の体長を  $(Y_1, Y_2, \dots, Y_n)$  とする。ただし、各親は、雌の子供は 1 個体しか生まないとする。親の体長の標本平均  $\mu_x$  と標本標準偏差  $\sigma_x$  で、親  $i$  の体長  $X_i$  を  $x_i = (X_i - \mu_x) / \sigma_x$  と正規化し、また子の体長の標本平均  $\mu_y$  と標本標準偏差  $\sigma_y$  で、子  $i$  の体長  $Y_i$  を  $y_i = (Y_i - \mu_y) / \sigma_y$  と正規化し、 $x$  に対して、対応する  $y$  をプロットしたものが左に示す図である。 $y$  の  $x$  に対する回帰分析を行ったところ、原点を通る直線が得られた。（50 点）



プロットしたものが左に示す図である。 $y$  の  $x$  に対する回帰分析を行ったところ、原点を通る直線が得られた。（50 点）

- (1) この回帰直線の回帰係数を、 $x_i, y_i, \mu_x, \mu_y, \sigma_x, \sigma_y, n$  を用いて表しなさい。
- (2) この回帰係数と、 $x$  と  $y$  の間の標本相関係数の関係を述べなさい。
- (3)  $y$  の  $x$  に対する回帰直線として適当なものは左図中の①、②、③の直線のいずれか。理由を述べて答えなさい。

- [2] ヒト、サル、ウシ、マウス、トリ間で相同なあるタンパク質のアミノ酸配列をアラインメントし、その系統樹を描くと下図のようになった。サルの配列がヒトではなくウシの配列に近縁であることがわかる。通常は、サルはヒトの配列と近縁になるはずだが、このような系統樹が得られた理由として考えられることを説明しなさい。考えられる理由が複数ある場合は、それら全てについて説明しなさい。（50 点）

