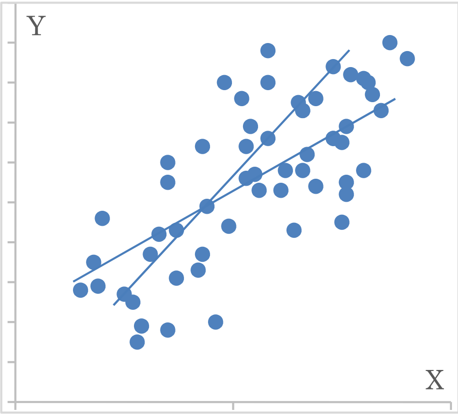
第５章　相互の関係

１．数量の相互の関係

算数と国語の得点、家庭で学習時間と算数の成績などの関係を調べるのに、どのように相関があるか求めます。たとえば、図の X と Y のように、得点が分布しているとき、X と Y の関係を調べる方法を次に説明します。



実際に得られた測定値 Y を Y＝aX+b の直線の式にします。しかし、Y には“ばらつき”があるため、Y をすべて結んでも直線にはなりません。

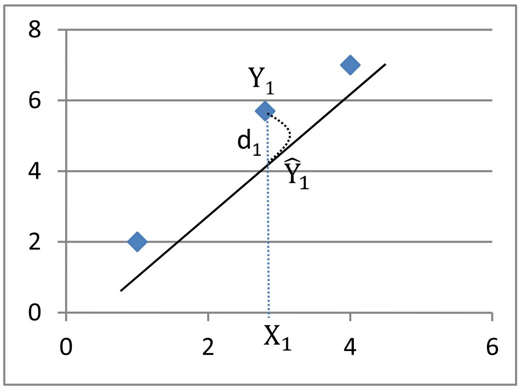
そこで、測定値 Y_1 と直線状の推定値 $\hat{Y}$ の距離を d とし、

$$d_1 = Y_1 - \hat{Y}_1 = Y_1 - aX_1 - b$$

Y_1 と $\hat{Y}_1$ の距離が最小となる値を d_1 とします。これが最小二乗法の基礎になります。

X と Y の二種類の測定値の関係をみてみましょう。

分布の回帰曲線は、X から見たとき、Y から見たときのそれぞれ二本の線になりますね。



回帰係数　Y＝*a*X+*b*　X＝*a’*Y+*b’*

$\hat{Y} = aX + b$　・・・①

$\hat{X} = a'Y + b$　・・・②

推定値　$\hat{Y}_1$、$\hat{Y}_2$・・・$\hat{X}_1$、$\hat{X}_2$・・・、

実際の観測値　Y_1、Y_2・・・X_1、X_2・・・

Y_1、Y_2・・・が分かれば、①より $\hat{Y} = aX_1 + b$

$d_1 = Y_1 - \hat{Y}_1 = Y_1 - aX_1 - b$

（$Y_1 = aX_1 + b$）

d の値には正負があるので、二乗和にします。

$Q_1 = {d_1}^2 + {d_2}^2 + {d_3}^2 + \cdots + {d_n}^2$

最小にする a を求めます。（最小二乗法）

この A について、X と Y を逆にした X=*a’*Y+*b’*の a’を同時に求めると次のようになります。

$$a = \frac{\frac{1}{N}\sum_{i=1}^{n}(Xi - \bar{X})(Yi - \bar{Y})}{\frac{1}{N}\sum_{i=1}^{n}(Xi - \bar{X})^2} \qquad a' = \frac{\frac{1}{N}\sum_{i=1}^{n}(Xi - \bar{X})(Yi - \bar{Y})}{\frac{1}{N}\sum_{i=1}^{n}(Yi - \bar{Y})^2}$$

r=ピアソン（K.Pearson）の偏差積率相関係数

$$r=\sqrt{aa'}=\frac{\frac{1}{N}\sum_{i=1}^{n}(Xi-\bar{X})(Yi-\bar{Y})}{\sqrt{\frac{1}{N}\sum_{i=1}^{n}(Xi-\bar{X})^2}\sqrt{\frac{1}{N}\sum_{i=1}^{n}(Yi-\bar{Y})^2}}$$

相関係数も表計算で処理できますので、式を覚える必要はありません。しかし、**「r の取りうる範囲」**については、きちんと覚えておく必要があります。

（使用例：二つの教科の試験結果の相関の有無を調べる）

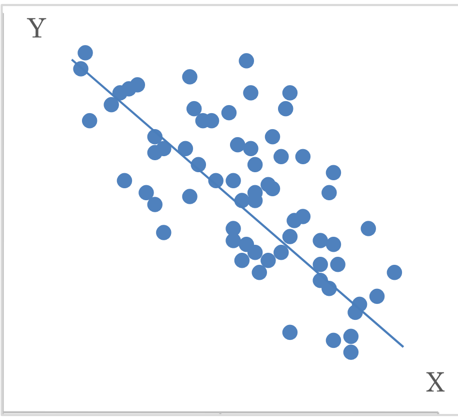
r の取りうる範囲　1～-1

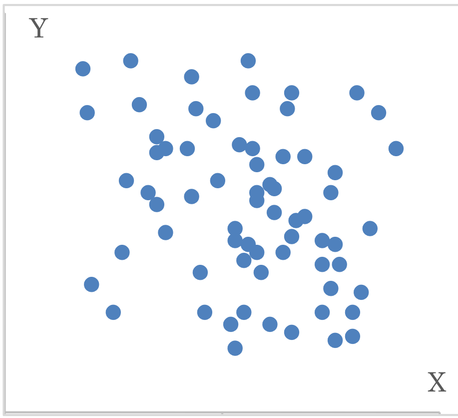
0.20～0.40	低い相関がある
0.40～0.70	かなり相関がある
0.70～1.00	高い相関がある

相関係数は、パソコンなどの表計算で処理ができますので、一般に計算する必要はありません。それで相関係数の意味、解釈をしっかり学修しておいてください。

また、これまでは、正の相関（x が大きくなれば y も大きくなる）について説明しましたが、負の相関もあります。また、図のように相関のない場合もあります。

その他、いろいろな相関があります。しかし、教育の場合、多くは正の相関が多いようです。





２． φ係数

2 つの問題の正誤（1、0）の相互関係を考えます。

数値（量）データでは、算数、国語に相関があるかないか、さらにどの程度の関係があるかは、相関係数 r で求められました。このとき r=0.7 であれば高い相関があるといわれています。

たとえば問題 A と問題 B の正誤（1、0）について次のような関係について検討します。

B / A	0	1	計	P
0				
1				
計				
P				

Ⓐ　0　0　1　1

↓

Ⓑ　0　1　0　1

また、$P(A)=$　　　　$P(\bar{A}、\bar{B})=$

$P(\bar{A})=$　　　　$P(\bar{A}、B)=$

$P(B)=$　　　　$P(A、\bar{B})=$

$P(\bar{B})=$　　　　$P(A、B)=$

などを求めて下さい。

$$\phi=\frac{P(\bar{A}、\bar{B})P(A、B)-P(\bar{A}、B)P(A、\bar{B})}{\sqrt{P(\bar{A})P(A)P(\bar{B})P(B)}}$$

で求められます。

このφ係数は、P(A)、P(B)によって取り得る範囲が決まります。

$P(A)+P(B)\leqq 1$ ならば

$$\phi\min=\sqrt{\left(\frac{P(A)}{1-P(A)}\right)\left(\frac{P(B)}{1-P(B)}\right)}\qquad \phi\max=\sqrt{\left(\frac{P(A)}{1-P(A)}\right)\left(\frac{1-P(B)}{P(B)}\right)}$$

$P(A)+P(B)>1$ ならば

$$\phi\min=\sqrt{\left(\frac{1-P(A)}{P(A)}\right)\left(\frac{1-P(B)}{P(B)}\right)}\qquad \phi\max=\sqrt{\left(\frac{P(A)}{1-P(A)}\right)\left(\frac{1-P(B)}{P(B)}\right)}$$

このため、このφが取り得る値の最大、最小を計算し、0.6 でも最大の値が 0.65 であれば関係があることを示しています。

φ係数の式については、統計の図書等を見て下さい。