学籍番号	氏　名	提出日

4．単回帰分析

4-0　前回講義の Tips

前回講義では，データの散らばり具合を表す偏差平方和，分散や標準偏差，また２変数の関係を表す相関係数を，Excel で数回のステップに分けて求めました．考え方を学ぶというう えでは必要なことですが，毎回，同じように計算をしていては時間がもったいないですね．Excel には便利な関数があって，先ほどの４つの値を簡単に計算することができます．

表4-1　便利な Excel 関数

	関数	簡単な使い方
偏差平方和	DEVSQ	“=DEVSQ(C4:C10)” と入力すると，C4 から C10 のセルにあるデータの偏差平方和が計算される．
分散	VARP	“=VARP C4:C10)” と入力すると，C4 から C10 のセルにあるデータの分散が計算される．
標準偏差	STDEVP	“=STDEVP(C4:C10)” と入力すると，C4 から C10 のセルにあるデータの標準偏差が計算される．
相関係数	CORREL	“=CORREL(C5:C28,D5:D28)” と入力すると，C5 から C28 のセルにあるデータと D5 から D28 のセルにあるデータの相関係数が計算される．

注）詳しい使い方は，Excel 画面の右上にある「質問を入力して下さい」に調べたい関数や求めたい指標を入力して検索してみよう．

相関係数の注意点①：２変数の相関関係が高くても（相関係数が１に近くても），その二つの変数の因果関係は分かりません．

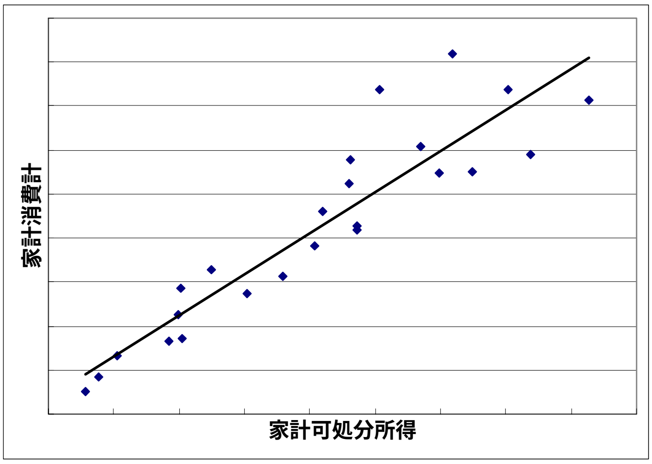
相関係数の注意点②：経済の時系列データは，それぞれ右上がりの直線トレンドに沿って増加するデータが多く，見せかけの相関が生み出されやすいので注意が必要です．たとえば，日本の高齢者数は，年々増えています．また，日本国内の高速道路の総延長も年々伸びています．この２つの時系列データには相関があります．しかし，高速道路が伸びたから高齢者が増えた，あるいは高齢者が増えたから高速道路が延びたとは言えないですね．このような相関を**見せかけの相関**と言います．

4-1　回帰分析とは

前回，コンビニM&Mの浦安地域マネージャーのあなたは，浦安市に新規出店して十分に売上げがあるのか（消費者がどのくらい購買意欲（＝支出）があるのか）を考えるため，家計可処分所得と家計消費支出の散布図を描き検討しました．また，その相関係数は 0.9951 と非常に大きく，可処分所得が増加すると家計消費支出が増加する傾向が強いことが分かりました．そこで，あなたは将来の売上げを考えるうえで参考となる将来の家計消費支出を回帰分析で予測することにしました．

回帰分析とは，２変数X，Yのデータがあるとき，YをXで定量的に説明する**回帰式**（あるいは，**回帰方程式**）とよばれる式を求めることを目的とします．前回の作業で描いた散布図は，家計可処分所得と家計消費支出の２種類のデータが平面上に点となって散らばっていました．簡単にいうと，回帰方程式は，その散らばっている無数の点を最もよく表現できる直線のことです．

図4-1　散布図と回帰方程式



学籍番号	氏　名	提出日

前々回，検討したように，消費支出は，「消費支出 C は所得 Y に依存するが，**限界消費性向**は１未満の正の数である」という仮説をたてて，次式のように線形関数に特定化しました．消費支出 C のように「説明される変数」は**被説明変数**（あるいは従属変数や内生変数）とよばれ，所得 Y のように「説明する変数」は**説明変数**（あるいは独立変数や外生変数）とよばれます．また，a は切片，b は傾きといい，所得 Y がが１単位の増加に対する消費支出 C の増分を示す限界性向を表しています．

回帰式：$C = a + bY$， $a \geq 0, 0 < b < 1$　　　・・・(1)

しかし，(1)式の消費関数には，定量的に把握することができないデータ（夏のすごしやすさや花粉症のひどさのこと）の影響が考慮されていません．そこで，i番目の観測データ（C_i, Y_i）では把握できないバラツキであるε_iを考慮したモデルが，次式の確率モデルでした．この方程式は**母回帰式**（あるいは，**母回帰方程式**）とよばれます．なお，aやbは**母回帰係数**とよばれます．

母回帰方程式：$C_i = a + bY_i + \varepsilon_i$， $a \geq 0, 0 < b < 1$　　　・・・(2)

説明変数Y_iや誤差項ε_iが，以下の４つの条件

条件①：Y_iは確率変数ではなく，確定した値をとる．
条件②：ε_iは確率変数で期待値０．$E(\varepsilon_i) = 0$，$i = 1, 2, \cdots, n$.
条件③：異なった誤差項は無相関．すなわち，$i \neq j$であれば，ε_iとε_jの共分散は０.
条件④：誤差項の分散が一定でσ^2.

を満たすとき，消費支出C_iの期待値は，次式のようになります．

$$E[C_i] = a + bY_i \qquad \cdots(3)$$

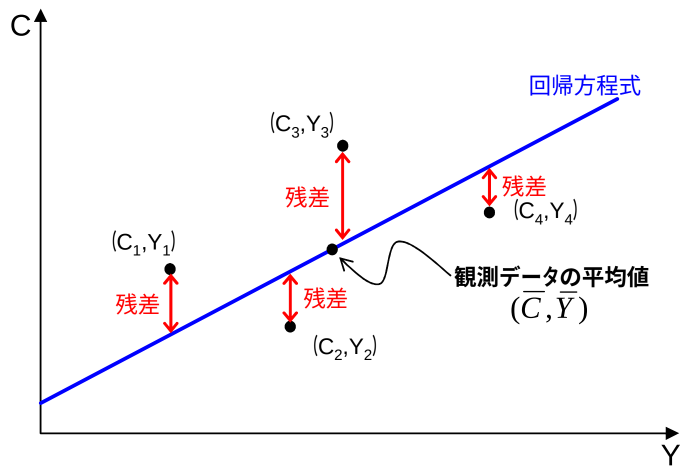
すなわち，説明変数や誤差項が①～④の条件を満たすと考えられるとき，それぞれの観測データで把握することができないバラツキ（誤差）をならした平均的な直線は，回帰式にほかならないということです．

4-2　最小二乗法

誤差項ε_iは，(2)式を変形した次式から，観測される消費支出C_iと回帰式（$a + bY_i$）から計算される**理論値**（**あてはまり値**）との差分であることが分かります．散布図では，下のイメージ図にある縦矢印部分で表され，**残差**とよばれます．

誤差項：$\varepsilon_i = C_i - (a + bY_i)$　　　・・・(4)

図４-２　回帰方程式と残差



回帰式は，この残差が最も小さくなるような母回帰係数 a，b として求められます．ただし，(4)式で表される残差は，プラスやマイナスの両方があり相殺されることがあるため，回帰式は残差を**二乗**したものの合計が**最も小さく**なるように求められます．この方法を**最小二乗法**といいます．

さて，この最小二乗法をつかって，母回帰係数をどうやって求めればいいのか？ みなさんが苦手な微分を使います^^; 講義時間が

学籍番号	氏　名	提出日

足りませんので，ここでは具体的な求め方のステップだけ紹介するにとどめておきます．詳しく知りたい方は，計量経済分析や多変量解析のテキストをじっくり読んで下さい．

STEP1：残差平方和を求める母回帰係数 a, b で偏微分したものを０とする（一階条件）．
STEP2：正規方程式（a と b に関する連立方程式）から a, b を算出する．

以上の STEP から母回帰係数の最小二乗推定値が次式のように計算されます．

母回帰係数：
$$b=\frac{C\text{と}Y\text{の偏差積和}}{Y\text{の偏差平方和}}=\frac{S_{CY}}{S_{YY}}=\frac{S_{CY}}{S_{YY}} \qquad \cdots(5)$$
$$a=\overline{C}-\overline{Y}\times b$$

ここで，母回帰係数 b の分母にある偏差平方和は前回学んだ方法で計算でき，分子にある偏差積和は次式で計算できます．なお $\overline{C},\overline{Y}$ は C,Y の平均値を表していおり，(5)式から分かるように，回帰式は必ず観測データの平均値 $(\overline{C},\overline{Y})$ を通ります．

偏差積和：
$$S_{CY}=\sum\left(C-C\text{の平均値}\right)\left(Y-Y\text{の平均値}\right) \qquad \cdots(6)$$

作業 1．**Ecoome06_4a.xls をＤＬして，****data_04a と名前を変えて各自のＵＦ（ユーザーフォルダー）の計量分析の中にセーブしてください．sheet name “家計消費支出（その３）”で，母回帰係数 a,b をステップ①～⑧の順番で求めなさい．****は学籍番号．（２点×８）**

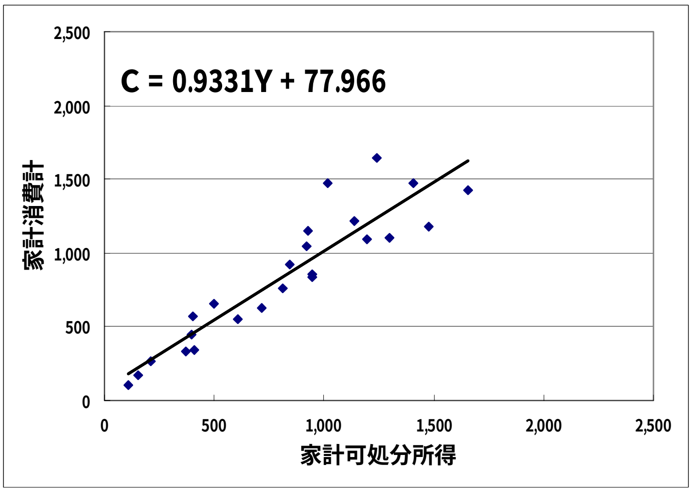
作業 2．******data_04a の sheet name “家計消費支出（その３）”において，作業１で推定した母回帰係数を使い，家計消費の理論値を求めなさい（⑨に記入）．また，家計消費支出と家計可処分所得の散布図にもともとのデータと理論値をプロットしなさい．****は学籍番号．（２点×１，５点×１）**

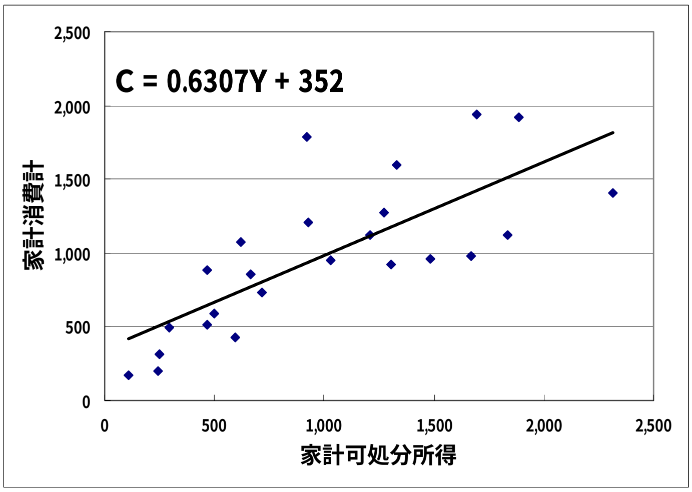
4-3　回帰式の精度

回帰式は，観測データさえあれば，最小二乗法の考え方で求めることができます．しかし，求められた回帰式が優れているかどうかはきちんと確認する必要があります．下図を見比べてみて下さい，何か違いがありませんか？

- 右図と比べると，左図のほうの点と回帰式がよりマッチしている．

図４-３　２種類の観測データと回帰式





左図のように，点(観測データ)と回帰式がマッチしているほど，回帰式の精度が高いといいます．しかし，回帰式の精度をグラフからマ

学籍番号	氏　名	提出日

ッチしているかどうか判断していては，グラフを見る人の主観が入ってしまいます．そこで，データの散らばり具合を表す指標として相関係数などを考えたように，回帰式の精度を表す指標としては**重相関係数**や**決定係数**（あるいは，**寄与率**）があります．

重相関係数は，消費支出の実測値（観測値）C と回帰式から計算される理論値 $\hat{C}$ の相関係数のことで，次式で計算できます．（一般的に，理論値やパラメータの推定値にはハット記号が使われます）

$$\text{重相関係数：}\frac{\left(C\text{と}\hat{C}\text{の偏差積和}\right)}{\sqrt{\left(C\text{の偏差平方和}\right)\left(\hat{C}\text{の偏差平方和}\right)}}=\frac{S_{C\hat{C}}}{\sqrt{S_{CC}\cdot S_{\hat{C}\hat{C}}}}\qquad\cdots(7)$$

なお，前回講義で学んだ相関係数の公式とは若干異なりますが（共分散が偏差積和へ，標準偏差が偏差平方和へ変わっています），各自，相関係数の公式から重相関係数の公式を導き出してみて下さい．

- 重相関係数は，観測値と理論値の相関係数である．このため，回帰式の精度が高いほど１に近づき，精度が悪くなるほど０に近づく．

回帰式の精度を表す２つ目の指標である決定係数は，重相関係数の二乗に等しい．（決定係数と重相関係数の関係の証明を省く，時間があったら挑戦してみよう）決定係数は消費支出 C_i の観測値平均まわりでの変動和（**総変動**という）と，理論値 $\hat{C}$ のその平均まわりでの変動和（**回帰変動**という）の比率に等しく，回帰式が観測値をどのくらい説明しているかを表している．総変動のうち，回帰変動で説明できないものは残差による変動になる．したがって，次式のように式変形できる．

$$\text{決定係数：}\left(\text{重相関係数}\right)^2=\frac{\left(\text{回帰変動}\right)}{\left(\text{総変動}\right)}=1-\frac{\left(\text{残差変動}\right)}{\left(\text{総変動}\right)}\qquad\cdots(8)$$

$$\text{総変動，残差変動：}\left(\text{総変動}\right)=\sum\left(C_i-\overline{C}\right)^2,\quad\left(\text{残差変動}\right)=\sum\left(C_i-\hat{C}_i\right)^2\quad\cdots(9)$$

- 決定係数は，観測値を理論値がどれだけ説明しているかを表し，回帰式の精度が高いほど１に近づき，精度が悪くなるほど０に近づく．
- 決定係数が１のとき，観測値は回帰式によって１００％説明されている．

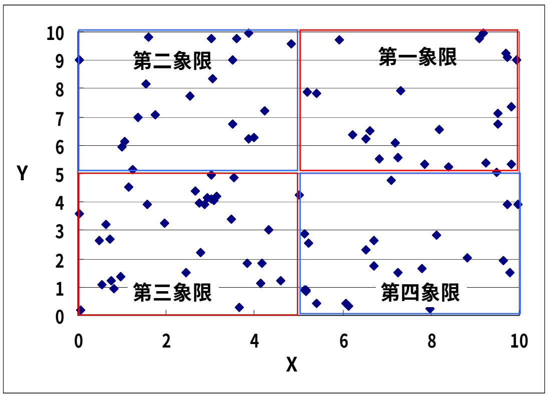
作業 3．******data_04a の sheet name “ 家計消費支出（その４） ”で，作業１で求めた母回帰係数 a,b を利用して，回帰式の決定係数をステップ①～⑧の順番で求めなさい．****は学籍番号．（２点×８）**

学籍番号	氏　名	提出日

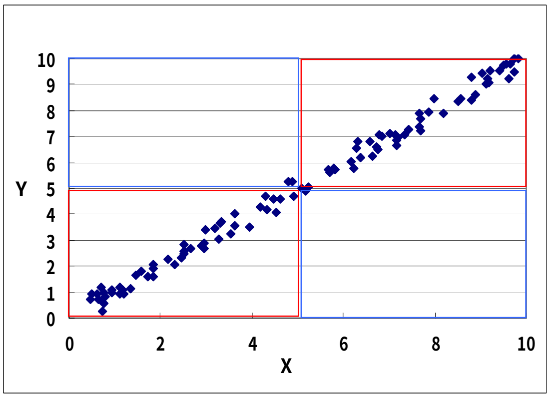
4-4　共分散と相関係数の関係

共分散は二つの変数の偏差の積です．その結果，平均値（$\bar{X},\bar{Y}$）を原点として，第一象限，第三象限のペアではプラスになり，第二象限，第四象限ではマイナスになります．例えば，相関係数 0.083 のように散らばっていると，プラスの値とマイナスの値が混ざっているため共分散の総和は小さくなりますし，0.996 のように右上がりの場合は第一，第三象限が多くなるため総和はプラスの値，下の-0.981 のような右下がりの場合は第二，第四象限のペアが多くなるため総和がマイナスの値になることが分かります．

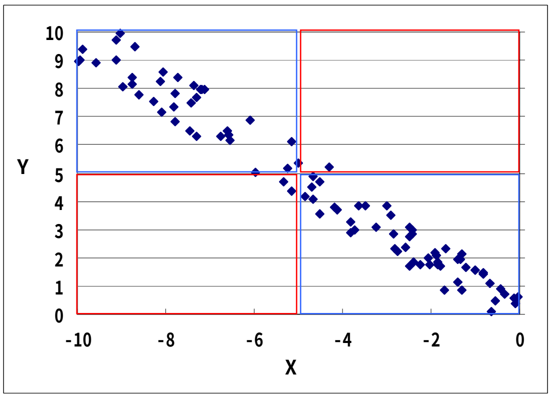
図４-４　共分散と相関係数の関係



相関係数＝０．０８３



相関係数＝０．９９６



相関係数＝－０．９８１