**3. 相関係数**

　　相関係数を求める作業は，2変数の直線関係の存在と強度を求めることにある。ここでは，おもに身長と体重の例で述べる。

1.　2変数の関係を大雑把に把握するために，まずは**散布図**を作成する。

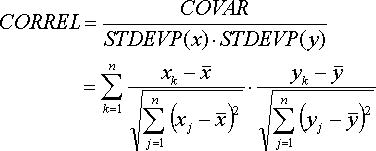
　直線関係がなくても曲線関係がある場合があり，そういう2変数関係の相関性を求めるには，スピアマンの順位相関係数の検定Spearman's correlation coefficient by rank testがあり，ここでおもに述べるピアソンの相関係数の検定Peason's correlation coefficient testを使わないことになる。ピアソンの検定は2変数が正規性を持つ場合に限られる。すでにこれまでに，身長と体重については，正規性を確かめており，この例ではピアソンの検定を使うことになる。

2.　散布図の作成によって，直線性がおよそ見られるので，**ピアソンの相関係数r**を求めることになる。式や図の作成が面倒なので，下記のサイトを使わさせていただく。 Excel 統計超入門　第4回　散布図，相関係数，回帰直線で測るデータの関連度　by 福島県立医科大学医学部数学講座

<https://www.fmu.ac.jp/home/mathema/lec/excel-static/statistics-04.htm>

https://www.fmu.ac.jp/home/mathema/lec/excel-static/image/image0402.gif

　上記の共分散covarianceは，式から見て取れるように，2変量(x,y)の各ペアについて，それぞれの平均からの差の積の総和を意味している。この共分散は，標本数で割ってはいるが，標本集団の数値の大小に依存しており，他の標本集団から求められる共分散との比較ができない。そこで，2変量(x,y)それぞれの標準偏差の積で割ることで，最大値の絶対値を1以内に留めることができる。その式を次に示す。



　上式の第2行目はその関係を示している。xの偏差平方和S(xx)，yの偏差平方和S(yy)，xとyの偏差積和S(xy)を使って，Σ(S(xy)/√(S(xx)・S(yy)))とも表現できる。covar=1/nΣS(xy)，stdevp(x)=√(1/n・S(xx))，stdevp(y)=√(1/n・S(yy))，なので，分母と分子の1/nは約分され，エクセルの関数を使った第1行が成り立つ。エクセルの関数の使用法： covar(配列1, 配列2)，correl(配列1, 配列2)。

　相関係数は2変量(x,y)の関数関係を示すものではない。独立変数x，従属変数yという関係は無い。相関係数の式を見てもわかるように，2変量(x,y)のx,yは入れ替えても同じ値を得ることができる。

　相関係数の符号が＋の時には正の相関関係，—の時には負の相関関係がある。相関関係の強さは絶対値|r|（rは上式のcorrel），または2乗値r^2（決定係数と呼ぶ）で評価する。

　2変量(x,y)いずれか，または両方に正規性が無い時や，直線性ではなくて曲線に沿って散布している場合には，rank関数を使って各変数の数値に代わって順位にして，ピアソンの相関係数を求めると，結果として，**スピアマンの順位相関係数rs**と同等になる。

3.　これまで求めてきた相関係数は標本集団のものであり，**母相関係数ρを推定**し，それを**検定**する必要がある。

3.1 無相関の検定

手順1　帰無仮説 H0: ρ=0 （母相関係数は0である）を設定し，これを否定できれば，標本から求めた相関係数rはゼロではなくて，有意になる。

手順2　検定統計量t値の計算をする。母相関係数ρが0の場合は，標本相関係数rを使って求めた次のt値は，自由度n（標本数）-2のt分布に従う。

検定統計量t値は，|r| √(n-2)/√(1-r^2)=abs(r)\*sqrt(n-2)/sqrt(1-r^2)で求めることができる。

手順3　これに対応するP値を求める。p値=tdist(t値,自由度，両側確率の場合は2)=tdist(t,n-2,2)。

手順4　有意水準α=0.05よりも小さい場合，帰無仮説（母集団は無相関）は棄却され，相関があることになる。

3.2 母相関係数の区間推定

　無相関の検定で母相関係数が0でないことが明らかになれば，母相関係数ρの区間推定を実施することになる。計算プロセスは省略して母相関係数の信頼区間上限と下限の式を示す。





　FISHERINV関数はフィッシャー変換の逆関数を返す。ここで，引数ZuとZlを求める必要がある。





　ここでは，zが必要になる。



　fisher(r)のrは標本相関係数だから，求めることができる。

以上から，次のように計算することになる。

手順1　z=fisher(r)

手順2　有意水準α=0.05とする。normsinv（normal standard inverse）関数は，累積標準正規分布の逆関数の値を返す。前掲の式を次に再掲する。





手順３　これで，ZuとZlが求まるので，下に再掲した式で，fisherinv(zu)とfisherinv(zl)を求めることができる。





以上