回帰分析（直線回帰）

２種類のデータの関係を調べる統計方法の１つとして「回帰分析」があります。相関分析は、２種類のデータの間に有意な関係があるかどうか、また、その関係はどの程度の強さがあるかを調べるものでした。これに対し、**回帰分析は、一方のデータからもう一方のデータを推定する計算式を求める分析方法です。**

従って、既に有意な直線関係があることが分かっているデータで分析を行わないと意味がありません。このため、通常は、先に相関分析で有意な直線関係があることを示した後に行う必要があります。

また、相関分析では、単純に２種類のデータの直線関係を調べるだけだったので、計算式のＸやＹにどちらの変数を入れても問題はありませんでした。しかし、回帰分析では、基準とするデータと推定するデータを明確する必要があるため、計算式のＸには必ず基準とする変数を、Ｙには推定する変数を入れなければなりません。

更に、**推定できる関係はあくまで直線関係だけ**です。曲線関係などを求める場合は、重回帰分析を行う必要があります。

(1) 回帰分析を行う際の注意

回帰分析を行う際には、その前に、下記のような点をチェックしておく必要があります。

1) 回帰分析より先に、調べる２種類のデータの間に有意な直線関係があるかどうかを調べる。

このために下記の操作を行います。

1)-(1) 散布図を表示して、直線関係があることを確認する。

1)-(2) 相関係数を求め、有意な直線関係があり、かつ、強い相関関係があることを確認する。

「強い相関関係」を示す相関係数の大きさは、回帰式を求める目的によって異なる。得られた相関係数が回帰式を求める目的に合致する程度の強さならば、回帰分析を行う意味がある。

一般的には、最低でも「0.7」以上は必要と思われる。また、精度の高い予測式として回帰式を求めるならば、最低でも「0.9」以上は必要となる。

2) 直線以外の関係がある場合は、重回帰分析に変更する。

この場合も、必ず事前に、散布図の作成と重相関分析を行い、有意な関係を示す曲線式があることを確認する。

3) どちらの変数が基準となる値(説明変数)で、どちらが予測する変数(予測変数)となるのかを明確にする。

回帰分析で回帰式を求める目的が何なのかが明確ならば、どちらの変数を元に、どちらの変数を予測したいのかは明らかなはずです。もし、これがはっきりしない場合は、もう一度、回帰分析を行う目的を見直しましょう。

4) どちらの変数も正規分布していることを確認する。

両方の変数が正規分布していない場合は、正確な回帰式が求められない。このため、可能ならば関数変換を行い、正規分布させることが望ましい。もし、正規分布するように変換可能な関数が見つからない場合は、回帰分析を中止することも検討が必要である。

なお、サンプル数が 100 を越える場合は、正規分布していなくても、比較的正確な回帰式を求められることがあるで、一度分析を行ってみるのも良い。

5) 散布図で、他のデータから飛び離れたデータが無いこと確認する。

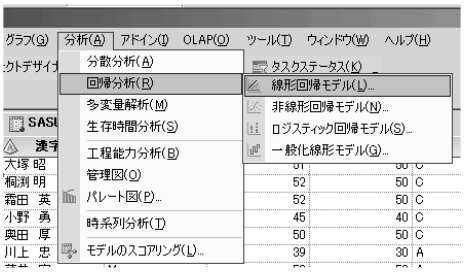
散布図を表示した際に、他のデータから飛び離れた少数のデータが存在すると、回帰分析や相関分析に影響を与え、不正確な結果に導かれる危険がある。

このため、もし、飛び離れた少数のデータが存在する場合は、それらのサンプルに何らかの問題が無いかを再検討する。(例えば、それらのデータのみ測定機器が異なっている、あるいは測定者の訓練が不十分であった等) もし、データが正確でない疑いがある要因が見つかれば、その要因を持っているサンプル全てを除外して、分析を行う。

なお、データが正確であることを疑う要因が見つからない場合は、分析を中止することも検討する。**間違っても、そのデータが飛び離れているという理由だけで、そのサンプルを除外してはならない。そのような行為は、ある意味でデータの捏造に準ずる行為である。**

(2) SAS での回帰分析を行う方法

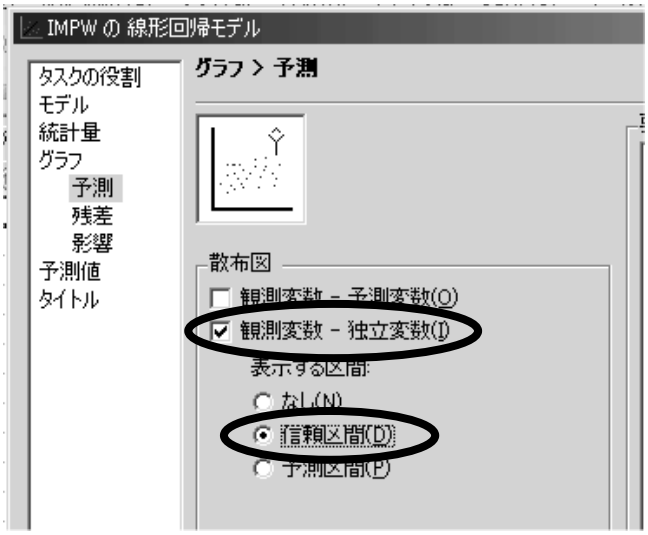
1) 上部メニュー内の「分析」をクリックし、その下に新たに表示されたメニュー内の「回帰分析」をクリックし、更にその右に表示されたメニュー内の「線形回帰モデル」をクリックする。



2) 新たに「線形回帰モデル」のダイアログが開くので、まず、「変数リスト」の下のリストから、予測の基準となる変数(説明変数)を選択してからクリックし、右隣りの右向き矢印ボタンをクリックし、新たに表示されたメニューから「説明変数」をクリックして、予測の基準となる変数(説明変数)を「タスクの役割」の下の「説明変数」の下に表示させる。

3) 次に、予測する変数(予測変数)を「変数リスト」の下のリストから選択してクリックし、右隣りの右向き矢印ボタンをクリックし、新たに表示されたメニューから「従属変数」をクリックして、予測する変数(予測変数)を「タスクの役割」の下の「従属変数」の下に表示させる。

4) 左端のリストから、「グラフ」の下の「予測」をクリックし、右に新たに表示されたものの「散布図」の下の「観測変数 － 独立変数」の前のボックスをクリックしてチェックを入れる。更に、その下に表示された「信頼区間」の前の丸をクリックして、マークを入れる。



5) ダイアログ最下部の「実行」ボタンをクリックし、結果を表示させる。

6) 表示された結果のうち、「パラメータ推定値」の表を見る。

回帰式を、

［予測する変数(予測変数)］＝ a ［予測の基準となる変数(説明変数)］ ＋ b

とした場合、「パラメータ推定値」の表の中で、［予測の基準となる変数(説明変数)］の名前の書かれた行の「パラメータ推定値」の値が回帰式の「ａ」となり、「Intercept」と書かれた行の「パラメータ推定値」の値が回帰式の「ｂ」となる。

パラメータ推定値

変数	ラベル	自由度	パラメータ推定値	標準誤差	t 値	Pr > |t|
Intercept	Intercept	1	-74.36346	6.97441	-10.66	<.0001
身長(cm)	身長(cm)	1	0.83938	0.04158	20.19	<.0001

(2) 回帰分析で言える結論の限界

1) 有意な関係があること、あるいは、因果関係があることを前提として分析しているので、これらの前提が否定された場合、分析の意味はまったく失われる。

2) 予測される変数(予測変数)のばらつきが、予測の基準となる変数(説明変数)の値により有意に差が有る場合、求められた回帰式は正確でなくなる。(例えば、説明変数の値が「1.0」の場合の予測変数のばらつきと、説明変数の値が「10.0」の場合の予測変数のばらつきが有意に異なる場合は、分析で得られた回帰式は信頼できないということ)

このような場合、「重み付き最小二乗法」により回帰式を求める必要がある。

(ここでは詳細は述べないので、興味のある人は、統計の本で調べること)