回帰式の検定

求めた回帰式に意味があるかどうかを判定するために，次のような回帰に関する**分散分析表**を作成して仮説検定を行う．

要因	平方和	自由度	分散	分散比 F	p 値
回帰	S_R	1	$V_R = S_R$	V_R/V_e	F の確率
残差	S_e	$n-2$	$V_e = S_e/(n-2)$		
計	$S(xy)$	$n-1$			

この検定における仮説は次のとおりである．

帰無仮説 H_0:　$\beta = 0$ (母回帰係数は 0 である．つまり，回帰式には意味が無い)
対立仮説 H_1:　$\beta \neq 0$ (母回帰係数は 0 ではない．つまり，回帰式には意味がある．)

p 値（有意確率）< 0.05 ならば，回帰式には意味があると判定する．

残差の検討

はずれ値の発見や回帰式が適当かどうかを判断するために，残差に関して次のことを検討する．

1. 残差が正規分布にしたがっているかどうか
2. 残差と説明変数は無関係（独立）かどうか
3. 残差と目的変数の予測値は無関係かどうか
4. 残差の時間的変化に癖があるかどうか
5. 残差が極めて大きいデータがないかどうか

以上の 1. から 5. に関して適切に判断するための具体的な方法として，それぞれ

1. 残差の正規確率プロットを行う
2. 残差と説明変数の散布図を作成する
3. 残差と目的変数の散布図を作成する
4. 時間順に残差を並べた時系列プロット（折れ線グラフ）を作成する
5. 標準化残差（残差を残差の標準偏差で割ったもの）を計算し，絶対値が 3 以上ならはずれ値と判断する．

残差の散布図

残差の散布図には，１．残差と説明変数の散布図，２．残差と予測値の散布図がある．どちらの散布図も何の傾向も見られないという状態が望ましく，何らかの傾向が見られる場合は回帰式が不適当であることを示している．例えば，残差が放物線のような形をしていれば，説明変数を 2 乗した項を回帰式の中に加えて再度行う必要がある．また，はずれ値が存在する場合も何らかの傾向が見られる散布図になることがある．

回帰式の精度

母回帰式の信頼区間

$$y = -2.1321 + 1.0915x$$

なる直線は，サンプルサイズ $n = 30$ のデータに基づいて求めたものであるから，サンプルが変わると，回帰式も変わることになる．そこで，真の回帰式（母回帰式という）の存在範囲を推定することを考える．これを母回帰式の区間推定という．いま，母回帰式を

$$y = \beta_0 + \beta_1 x$$

とすると，この式の 95%信頼区間は,

$$(b_0 + b_1 x_0) + t(n-2, 0.05)\sqrt{\left\{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S(xx)}\right\} V_e}$$

となる．ここで，$t(n-2, 0.05)$ は自由度 $n-2$ の t 分布における 5%点，$\bar{x}$ は説明変数 x の平均値である．なお，信頼区間の幅が x の値によって異なることに注意したい．

個々のデータの予測区間

説明変数 x の値を x_0 に指定したときに，将来実現するであろう目的変数 y の値の信頼区間を予測区間という．信頼率 95%の予測区間は,

$$(b_0 + b_1 x_0) + t(n-2, 0.05)\sqrt{\left\{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S(xx)}\right\} V_e}$$