

回帰分析について

—表計算ソフトを使って—

ビジネス用の表計算ソフト（ここでは Excel を試してみる）には統計処理機能があり、回帰分析もその一つである。本稿では、この機能を使った数学教育を考えてみる。新学習指導要領では「情報」や「総合的な学習の時間」が目玉として新設され、パソコンが十分に使える環境になる。有名な数式処理ソフトウェア Mathematica や Maple 等には確かに見劣りする。でも相当の部分で代替することができるであろうと著者は考えたが、実際に実験してみると相当の差が存在する。この報告は Excel の実力と Mathematica の実力の比較である。元々、ビジネスソフトと科学技術計算用のソフトと比較するのが無謀かもしれないと思うが、そのソフトの限界を知り使えば意味があると思う。

また、計算機の結果を盲信してはいけないといけなことの指摘でもある。多くのアプリケーションソフトでは、その精度やプログラムの内容はブラックボックスであり、その結果を吟味することは余りなく皆が信頼しているように見受けられる。その根本には数値解析の誤差解析の問題があり、その重要性を指摘するものである。今後、あらゆる場面にコンピュータが登場してくるものと思われるが、その結果を吟味したり、考察することはコンピュータ利用の教育には必要であると思われる。

例として、小林道正著：文科系に生かす微積分（BLUE BACKS、講談社）の p. 23 では原油輸入量の変化のデータを採り上げる。その本では Mathematica を用いてグラフとその関数式を求めている。この本の結果を以下に記述しておく。

●小林道正氏の本の結果より（Mathematica による結果）

原油輸入量の変化（単位 100 万 kl）は以下の通りである。ここに、関数値の値は小林道正氏の本の Mathematica による結果。

Mathematica による近似式は $y = 220155 - 7296.9x + 80.5152x^2 - 0.295454x^3$ である。

その近似式を用いて関数値は次の通りである。

年	原油輸入量	関数値(Mathematica)
84	213	214
85	196	195
86	189	187
87	185	187
88	192	194
89	205	206
90	225	221
91	236	238
92	247	254

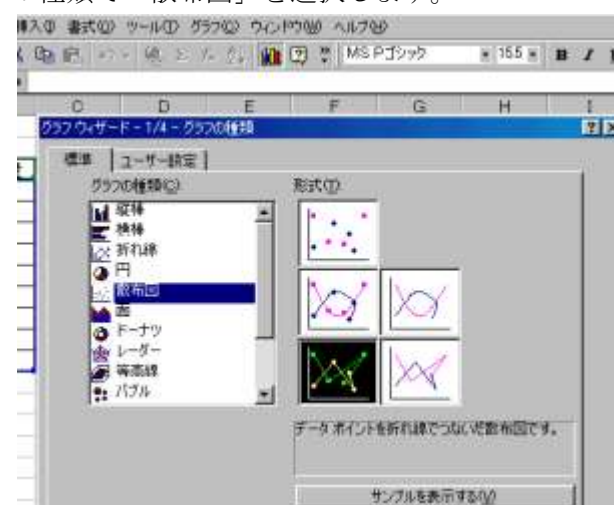


年	原油輸入量
1984	213
1985	196
1986	189
1987	185
1988	192
1989	205
1990	225
1991	236
1992	247

表を範囲選択
する

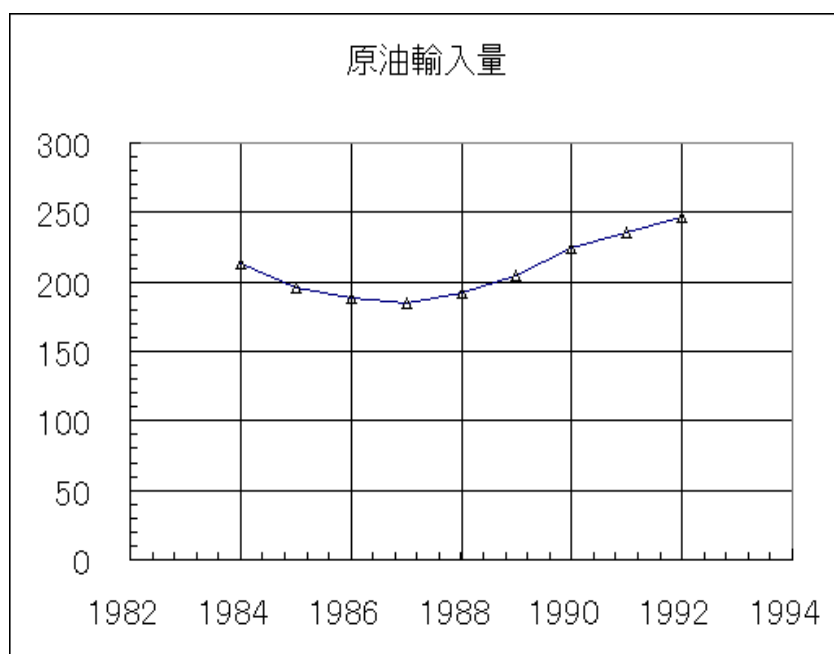
●Excel によるグラフの作成

上記の表の年と原油輸入量の列を範囲選択して、グラフウィザードのボタンをクリックします。そして、グラフの種類で「散布図」を選択します。



グラフの種類で「散布図」を選択

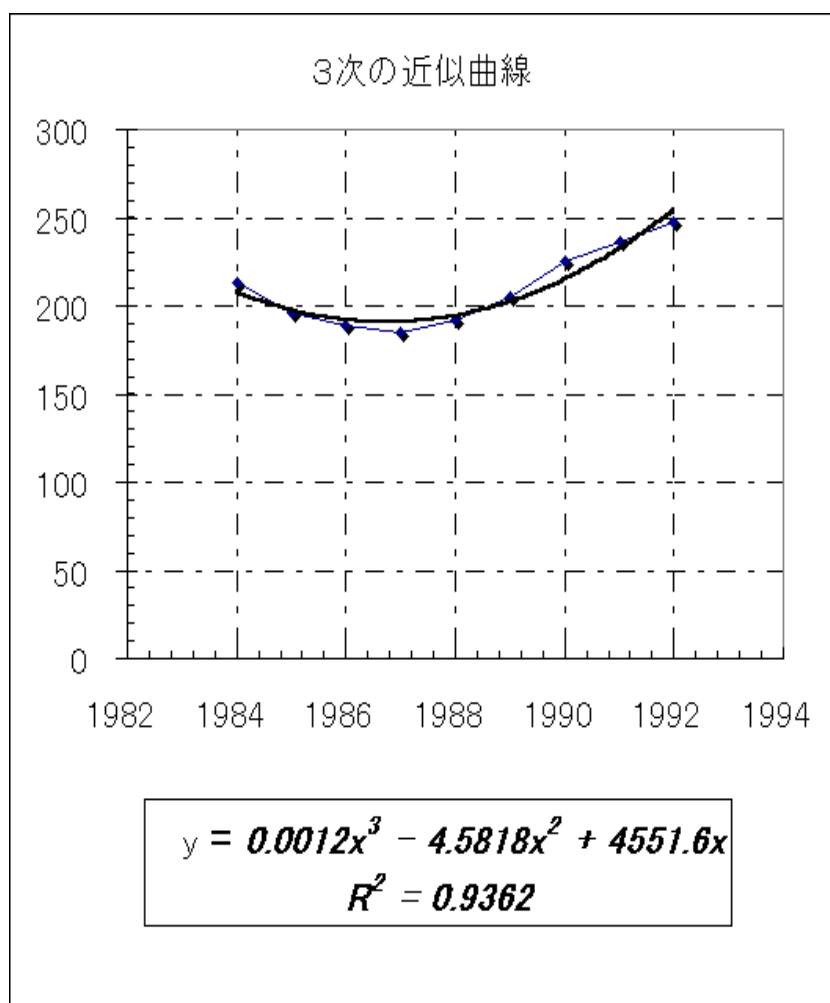
年	原油輸入量
1984	213
1985	196
1986	189
1987	185
1988	192
1989	205
1990	225
1991	236
1992	247



●Excel の回帰分析機能を使っての近似曲線の作成

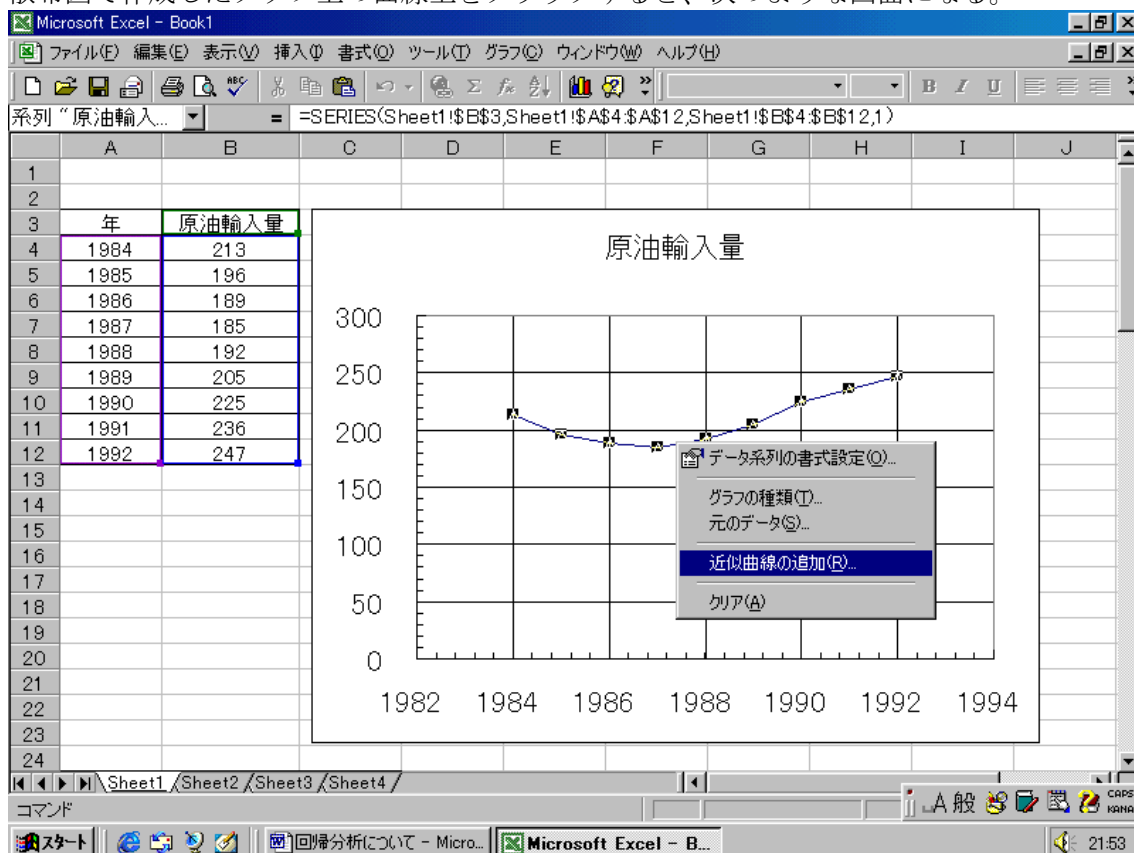
年	データ	関数値
1984	213	366674.9
1985	196	367219
1986	189	367768.3
1987	185	368322.6
1988	192	368882.1
1989	205	369446.8
1990	225	370016.6
1991	236	370591.6
1992	247	371171.8

ここに、上記の関数値は Excel の近似曲線から計算したものである。Mathematica によるものとはかなり違っている。Mathematica の方が精度がかなり優れている。



● 近似曲線の追加の仕方

散布図で作成したグラフ上の曲線上をクリックすると、次のような画面になる。



次の画面は多項式近似を選び、次数3を設定したところです。

The '近似曲線の追加' dialog box is shown with the '種類' (Type) tab selected. Under '近似または回帰の種類' (Type of trendline or regression), the '多項式近似(P)' (Polynomial trendline) option is selected. The '次数(D):' (Degree) is set to 3. The '区間(E):' (Interval) is set to 2. The '追加対象の系列(S):' (Series to add trendline to) list contains '原油輸入量'. The 'オプション' (Options) tab is also visible but not selected. The 'OK' and 'キャンセル' (Cancel) buttons are at the bottom.

さらに、近似曲線の方程式を求めるために、オプションのタブをクリックする。

そうすると、オプション画面になる。切片にチェックをいれ、Mathematica の結果を参考にして近似曲線 $y = 220155 - 7296.9x + 80.5152x^2 - 0.295454x^3$ の定数項の値 220155 を代入したところです。

近似曲線の書式設定

パターン | 種類 | オプション

近似曲線名

☒ 自動(A): 多項式 (原油輸入量)

☐ 指定(Q):

予測

前方補外(F): 0 単位

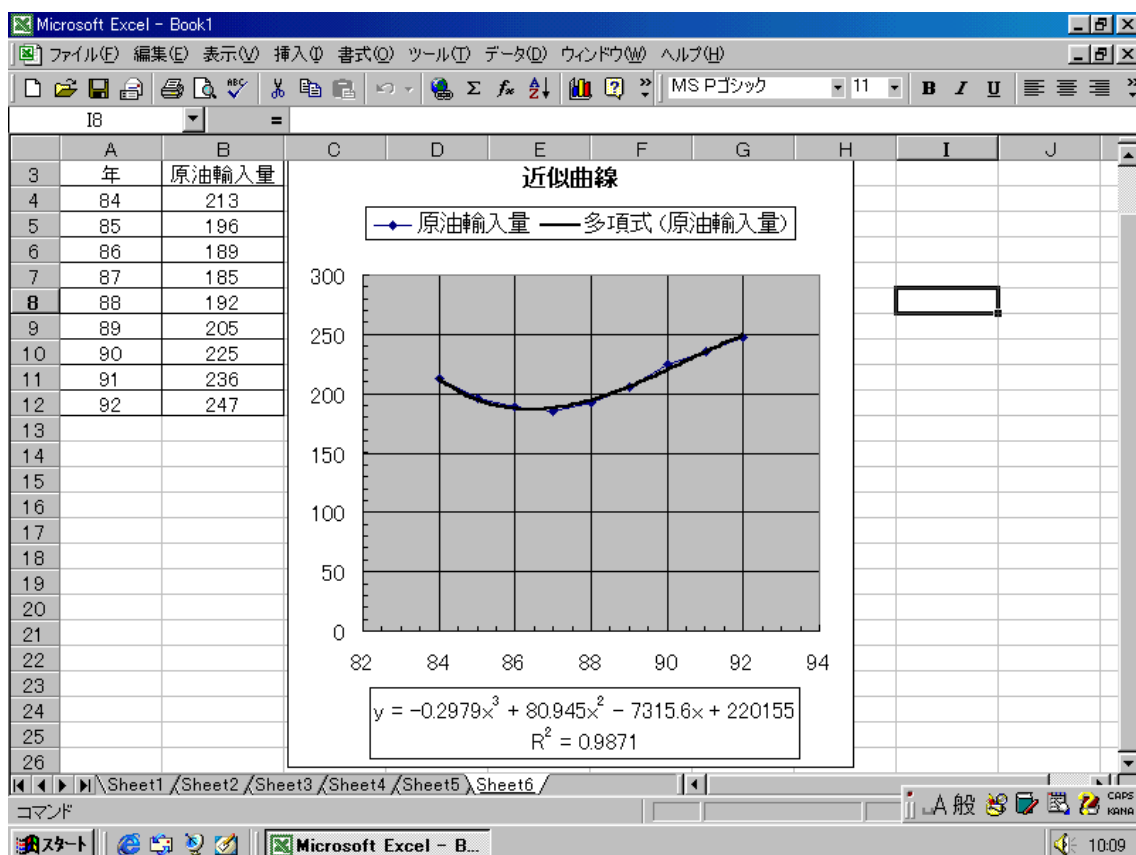
後方補外(B): 0 単位

☒ 切片(S) = 220155

☒ グラフに数式を表示する(E)

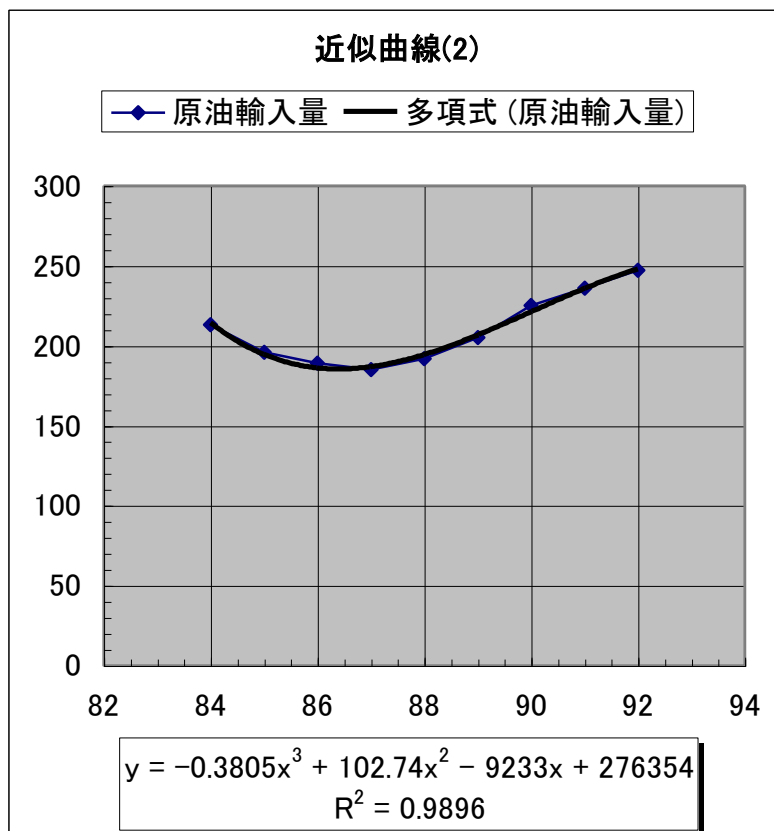
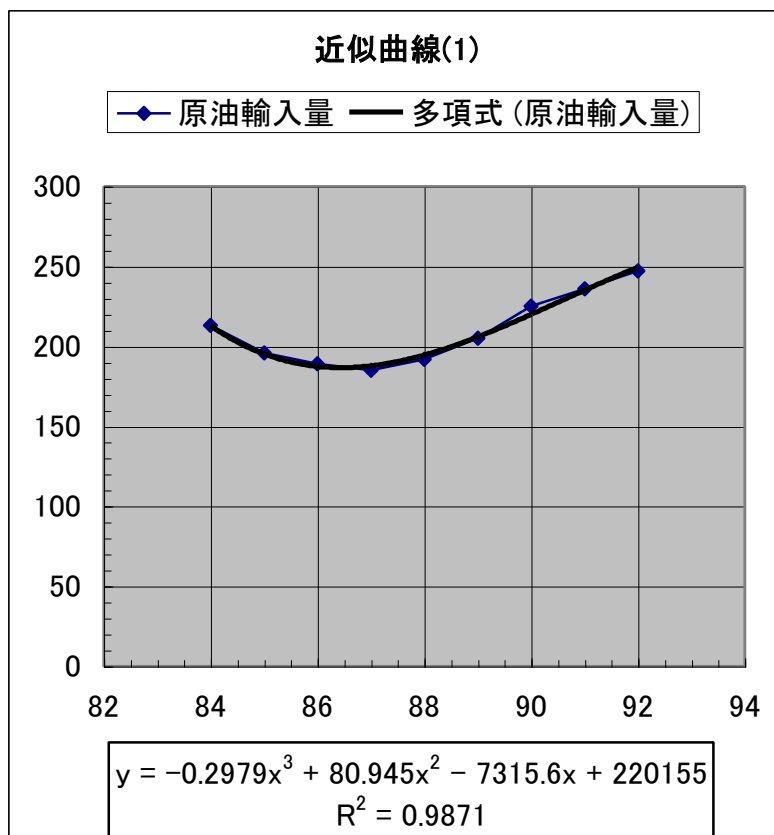
☒ グラフに R-2 乗値を表示する(R)

OK キャンセル



次に、切片にチェックをいれずに自動で近似曲線を作成してみる。

●最終結果（拡大図）



＊近似曲線の信頼性； [R-2 乗値](#)が 1 またはその近似値であるとき、近似曲線の信頼性が最も高くなります。近似曲線をデータに適合させると、その R-2 乗値が自動的に計算されます。必要に応じて、[グラフに R-2 乗値を表示](#)できます。

＊R-2 乗値；

$$R^2 = 1 - \frac{SSE}{SST} \quad \text{ただし、} SSE = \sum (Y_i - \hat{Y}_i)^2 \quad \text{および} \quad SST = \left(\sum Y_i^2 \right) - \frac{\left(\sum Y_i \right)^2}{n}.$$

● まとめ

近似曲線（１）、（２）の R-2 乗値を見比べてみると、Excel の方が Mathematica より優れているように見える。