



- [「Tiny Q」とは？](#)
- [「対話型変数選択法」とは？](#)
- [Tiny Q for Web の操作方法](#)
- [注意事項](#)
- [参考文献](#)
- [用語の説明](#)
- [謝辞](#)



## Tiny Q とは？

Tiny Q は「[対話型変数選択法](#)」による重回帰分析プログラムです。本プログラムは、多変量解析プログラム「まるば」を参考にして開発されています。「まるば」は、[小林龍一立教大学名誉教授](#)によって開発されたもので、重回帰分析・主成分分析・数量化1類・数量化2類・数量化3類・数量化4類などの多変量解析ができるという本格的なプログラムです。これに対して Tiny Q は「まるば」の「対話型変数選択法」に基づく重回帰分析のみを実現するプログラムです。Tiny Q はMacOSX版(Cocoa-Java)、Windows版（配布再開）、Pure Java版がありましたが、このページでは Web版について説明しています。

[トップへ](#)



## 「対話型変数選択法」とは？

Tiny Q を使えば、重回帰分析における説明変数の選択を対話的に行うことができます。以下、変数を選択するときの手順を簡単に説明します。

まず、それぞれの説明変数について「[変数選択のためのF-値](#)」を計算します。このF-値が最も大きな変数が、計算上最も説明力のある変数です。あなたは、分析の理論的含意を考慮しながら、F-値の高い変数を順番に説明変数として取り込んでいきます。大まかな目安としてF-値が2より大きいものを説明変数に取り込みます。最終的にF-値が2以上のものがなくなれば計算は終了します。

[トップへ](#)



## Tiny Q の操作方法

ここでは Web版を使った分析を紹介します（他のバージョンでも基本的な使い方は変わりません）。

はじめに、分析内容を簡単に説明しておきましょう。ここで取り上げる例は、企業規模や勤続年数を説明変数とした大卒男女の賃金構造の分析です。データは『賃金センサス』（2001年版）を使います。なお、この事例に関しては、蓑谷千凰彦『計量経済学 第3版』1997年（スタンダード経済学シリーズ、東洋経済新報社）を参考にしました。

データ・ファイルの構造は以下のとおりです。データの1行目にはデータ・ラベルとする必要があります。

sex	year	scale1	scale2	wage
1	0	1	0	2744.88
1	1.5	1	0	3456
1	3.5	1	0	3995.64
...	...	...	...	...

この他に性別×勤続年数（sex\*year）、規模1×勤続年数（scale1\*year）、規模2×勤続年数（scale2\*year）を加えています。

### <注>

ここで、賃金は第1巻第2表「年齢階級、勤続年数階級別所定内給与および年間賞与その他特別給与額」から、「所定内給与額」×12+「年間賞与等」としました。

また、勤続年数（year）、性別（sex）、企業規模（scale）は、以下のようにコーディングします。

year :1-2年=1.5, 3-4年=3.5, 5-9年=7,...

sex :male=1, female=0

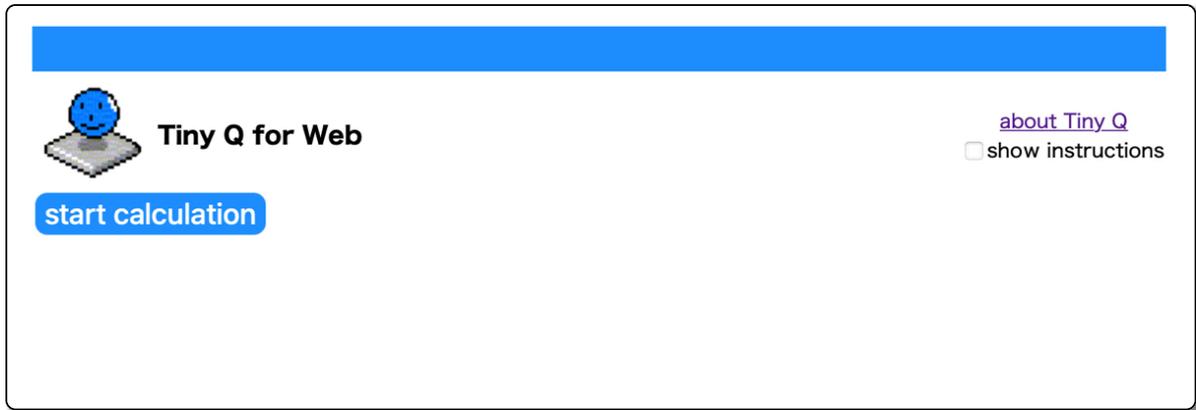
scale1: 1000人以上=1

scale2: 100-999人=1

## (1) TinyQ の起動

まずソフトウェアを起動します。Web版なので [Tiny Q のWebサイト](#) を開きます。ページを保存すればネットに繋がっていなくても動作するアプリケーションとして実行できます (Mac の Safari の場合、「別名で保存」をクリックし、Webアーカイブ形式で保存すると、関連ファイルも一緒に保存されます)。

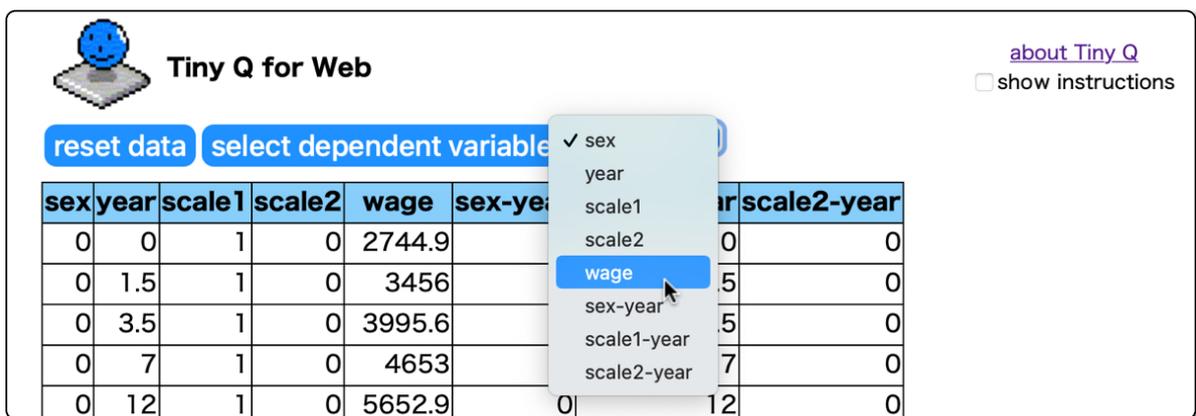
画面右の「show instructions」をチェックすると操作方法に関する簡単な説明を表示します。



## (2) データをインポートする

次にデータをインポートします。クリップボードからデータをインポートするか、CSV形式のデータをファイルからインポートすることができます。

クリップボードからインポートする場合は、[start calculation] ボタンをクリックするとテキストエリアが表示されるので、ここに Excel などからコピーしたデータをペーストし、[import data] ボタンをクリックします。CSV ファイルからインポートする場合は [ファイルを選択] ボタンをクリックし、ファイルを選択するとテキストエリアにデータが入力されるので、同様に [import data] ボタンをクリックします。すると、被説明変数を選択するためのセレクトボックスが表示されるので、wage を選び [select dependent variable] をクリックします。



### (3) 説明変数の選択

被説明変数を選択すると被説明変数 (dependent variable) が wage と表示され、その下には各変数のF-値が表示されます。F-値が最も大きい変数が最も説明力のある変数です。このF値を参考にして、回帰式に説明変数を取り込んでいきます。変数を取り込むと、F-値がマイナスになります。最初は回帰式に定数項だけが取り込まれた状態になっています。

F-値が最も大きいのが勤続年数 (year) の 351.395... なので、この変数をモデルに取り込みます。year の行をクリックすると変数が回帰式に取り込まれ、F-値が再計算されます。同様に、説明力の高い変数を選択していきます。変数を選択するかどうかの基準は、F 値の絶対値が 2 より大きいかどうかを目安とします。一度取り込んだ変数 (F-値がマイナスの変数) をクリックすると、その変数を回帰式から取り除くことができます。

**reset data** **end calculation**

dependent variable : wage

variables	F-value
const.	-337.378711
sex	13.269969
year	-351.395012
scale1	15.187963
scale2	0.272359
sex-year	7.112432
scale1-year	33.634802
scale2-year	0.212004

### (4) 結果の表示

性別 (sex)、勤続年数 (year)、規模1×勤続 (scale1-year)、規模2×勤続 (scale2-year) の4つの変数を選択した段階で、F-値が2を上回る変数がなくなりました。これで分析は終わりです。

[end calculation] ボタンをクリックして、分析結果を表示してみましょう。決定係数 ( $R^2$ )、自由度修正済み決定係数 (adjusted  $R^2$ )、説明変数の行列式の値 (det.)、各変数の偏回帰係数 (coefficient) と標準偏回帰係数 (std. coefficient) t 値が表示されます。

**reset data** **copy to clipboard**

dependent variable : wage

$R^2$ : 0.973697  
adjusted  $R^2$ : 0.971550  
det. : 0.523012 > 0.0001

variables	coefficient	std. coefficient	t-value
const.	3029.828322		29.672958
sex	765.177778	0.163143	7.041500
year	152.362603	0.711620	22.957278
scale1-year	90.540671	0.347095	11.856028
scale2-year	51.864432	0.198826	6.791491

重決定係数 (R<sup>2</sup>)            0.973697  
自由度調整済みR<sup>2</sup>            0.971550  
説明変数の行列式 (det.) 0.523012>0.0001

wage = 3029.82832 + 765.17778 \* sex + 152.36260 \* year + 90.54067 \* scale1-year +  
51.86443 \* scale2-year

[copy to clipboard] をクリックすると、計算結果がクリップボードにコピーされるので、Excel 等に貼り付けることができます。

---

## サンプル・データ

このサイトで使用しているデータは以下からダウンロードすることができます。

[ダウンロード](#)  
[wageData01.csv\(1.4KB\)](#)

[トップへ](#)

---

## 注意事項

Tiny Q は「まるば」がもつ多様な機能のうち、ほんの1部を実現するに過ぎません。より高度な分析をしたい方は「まるば」をご参照下さい。また、統計学の基本やプログラムについての詳細な解説は省略しています。詳しくは[参考文献](#)をご覧ください。

[トップへ](#)

---

## 参考文献

- 小林龍一『相関・回帰分析入門』日科技連出版社、1972年
- 小林龍一『パソコンによる統計解析』培風社、1983年
- 小林龍一『パソコンによる多変量解析』培風社、1984年
- 小林龍一『多変量解析プログラム まるば』1996年（自費出版）

[トップへ](#)



## 用語の説明

- 【CSV形式のデータ】

TinyQで利用可能なデータファイルは、データ要素がカンマで区切られた形式、いわゆるCSV (Comma Separated Value) 形式に限られる。CSV形式のファイルは一般的な表計算ソフトで利用可能である。第1行目はデータラベルを入力しておく必要がある。

- 【変数選択用のF値】

回帰分析において、各変数の説明力は、一般に編回帰係数をその標準偏差で割ったt値で判断される。「まるば」やTiny Qにおける変数選択用のF値は、このt値の2乗である。厳密には、F値の絶対値を、自由度1、 $n-k-1$ のF分布の上側5%点と比較して検定する( $n$ はデータ数、 $k$ は変数の数)。しかし、通常はF値が2以上であれば変数に説明力があると考えて変数を取込む。このプログラムでは取込んだ変数のF値はマイナスの値になっていることに注意。<

- 【多重共線性】

多重共線性とは、変数間の相関関係が非常に高い状態を指す。説明変数間に多重共線性が存在すると、偏相関係数の値が不安定になったり、その統計的有意性が低下したりする。Tiny Qでは相関行列の行列式を計算しているが、この値が $0.1^{(k-1)}$ より小さい場合は多重共線性の疑いがある。

- 【標準偏回帰係数】

編回帰係数を標準偏差で修正したもので、この値が大きい変数ほど被説明変数Yに対して影響力が大きい。

[トップへ](#)



## 謝辞

本プログラムの開発にあたり、「まるば」の開発者である小林龍一先生にアドバイスをいただきました。心から感謝いたします。

また、東京情報大学助教授（当時）の内田治先生には、本プログラムの誤りを指摘していただいたうえ、有益なコメントをいただきました。この場をかりてお礼申し上げます。

[トップへ](#)