

演習 1 回目の答えのワークシートの解説
< 回帰の実行 >

[J6~K10 までの四角]

=LINEST(C6:C189,D6:D189,TRUE,TRUE)

- ・ 最小二乗法による線形回帰を行う関数
- ・ =LINEST(被説明変数のデータ範囲,説明変数のデータ範囲,回帰式に定数項があるかどうか,TRUE)という形で使用。
- ・ 配列数式 = 複数のセル (データの格納場所) に計算式の結果を記憶させるので、まず 5 行 (定数項を含めた説明変数の数) 列の範囲をマウスでドラッグしマウスをそのままにしておいて、上の数式入力エリアに上記の数式を入力する。このとき、C6:C189 や D6:D189 などデータ範囲を指定する部分は、カーソル (縦線 |) を入力エリアに置いたまま、データ範囲のセルをドラッグして入力しても良い。このばあい、数式入力エリアで範囲指定ができれば、それ以上マウスをいじらず、数式入力エリアにキーボードから、カンマ(,)を入力することが肝心である。数式を完全に入力し終わったら、Ctrl キーと Shift キーを押したまま Enter キーを押す。
- ・ 出力は、二行目までは、各列に係数推定値、その標準誤差である。列の順番は、第 1 の説明変数、第 2 の説明変数、・・・、最後の説明変数、定数項の順である。この例では、第 1 列は限界消費性向を示す係数、第 2 列は定数項である。3 行目は、左から決定係数、残差標準偏差であり、のこりは #N/A(未定値)がはいる。4 行目は、F 統計量と、残差自由度 (n - K) である。5 行目は、ESS と RSS である。

< 係数推定値に関する統計量と信頼区間 >

[K15~L20 までの四角]

K15,L15: 推定値そのものなので、J6 と K6 から値を持ってくる。ただし、配列数式は複数のセルが 1 セットになっているので、一つのセルだけコピーすることはできない。したがって、J6 と K6 を参照することが必要なので、=J6 と=K6 という式になっている。入力の仕方は、まず K15 に=J6 と入力して、オートフィルを使う。オートフィルについては講義資料を参照してほしい。要は、右下の四角を横にマウスでドラッグすると、L15 に=K6 が入力される。ここで、なぜ=J6 ではないかという、K15 と L15 のセルとしての位置関係が、そのセルの中身である数式にも反映されるからである。つまり、L15 は K15 の右隣にあるので、L15 にある=J6 の式は右隣の K15 の中では、J6 の右隣の K6 を参照する、すなわち、=K6 となるのである。

K16,L16: それぞれの推定値の標準誤差。やはり、J7 と K7 から値を持ってくる。したがって、=J7 と=K7 という式を使う。入力にはオートフィルを使おう。

K17,L17: それぞれの推定値の t 統計量を計算する。一般的な t 統計量は係数 = 0 に対して計算されるので、係数推定値 / 標準誤差で計算される。したがって、上二つのセルの値の比をとればよい。結果として=K15/K16 と=J15/J16 をそれぞれ

のセルの数式として入力すれば、結果的に K15/K16 と J15/J16 の値が得られる。また、=J6/J7、=K6/K7 でもよい。入力にはオートフィルを使おう。なぜそれがうまくいくか考えてみよう。

K18,L18: それぞれの t 統計量に対する p 値。これは、いまのところこういう式を入力すれば得られると丸飲みしてほしい。=TDIST(t 統計量, 残差自由度, 2)であるが、K18 については、t 統計量は K17 にあり、残差自由度の値は K9 があるのでそれをいれてやればいい。すなわち、=TDIST(K17, K9, 2)である。ただし、定数項に関する p 値を計算する場合にオートフィルを使うとすると、L18 には=TDIST(L17, J9, 2)という数式が入ってしまう。J9 には残差自由度は入っていないので、この数式のままで問題が発生する。そこで、絶対参照を使うのである。具体的には、K9 と書く代わりに\$K\$9 と書くのである、こうするとオートフィルの際に、数式が入力されたセルの位置関係に応じて参照するセルが変わることがない。従って、K18 については、=TDIST(K17, \$K\$9, 2)と入力し、オートフィルを使用する。

K19~L20: それぞれの係数の値の 95%信頼区間を示す。上限は=係数推定値+標準誤差*TINV(0.05,残差自由度)、下限は=係数推定値 - 標準誤差*TINV(0.05,残差自由度)である。これも今のところ丸飲みしてほしい。絶対参照を使っているが、なぜこのような絶対参照を使用しているかは、めいめい考えてほしい。

< 回帰統計 >

K24: 決定係数である。決定係数は LINEST 関数の結果得られるので、それを参照する。具体的には、=J8 となる。

K25: 相関係数である。相関係数は、決定係数の平方根である。平方根を求める関数は=sqrt()であるから、=sqrt(決定係数)で計算できる。従って、=sqrt(K24)である。もちろん、=sqrt(J8)でもよい。

K26: 自由度修正済み決定係数 $1 - \frac{RSS / 残差自由度}{TSS / 全自由度} = 1 - \frac{RSS / (n - K)}{TSS / (n - 1)}$ であるが、RSS は K10、TSS は ESS+RSS なので、J10+K10、残差自由度は K9、全自由度は(184 - 1)である。したがって、= 1 - (K10/K9)/((J10+K10)/(184-1))で計算できる。

K27: 残差標準偏差。K8 に出力されていたから、=K8 となる。

< 分散分析 >

K31~K33: ESS, RSS, TSS が縦に並ぶ。ESS と RSS の値は、LINEST()関数の結果にあるので、それを取ってくる。従って、K31 は=J10、K32 は=K10 である。TSS は、ESS+RSS なので、=J10+K10、または、=K31+K32 である。

L31~L33: 回帰分散、残差分散、総分散を示す。回帰分散は ESS/(K-1)であるので、L31 には=J10/(2-1)が入るが、これは、=K31/(2-1)でもよい。残差分散は、RSS/残差自由度であるから、=K32/K 9 となるが、=K10/K9 でもよいし、残差標準

偏差の二乗であるから、 $=K8^2$ でもよい。ちなみに、 X^k は X の k 乗を示す。
 総分散は $TSS/(n-1)$ であるから、 $=K33/(184-1)$ である。

M31: F 統計量 .
$$\frac{\text{回帰分散}}{\text{残差分散}} = \frac{ESS/\text{回帰自由度}}{RSS/\text{残差自由度}} = \frac{ESS/(K-1)}{RSS/(n-K)}$$
 で計算

できるので、 $=L31/L32$ である。解答は、2, 3 番目の式を使って計算している。
 N31: F 統計量の P 値。これも丸飲み。 $=FDIST(M31,1,K9)$ となる。一般には、 $=FDIST(F$
 統計量, 回帰自由度, 残差自由度) となる。

< 予測値, 残差, 標準残差 >


E6~E189: 予測値は、限界消費性向の係数推定値 × 所得 + 定数項の推定値で計算できるので、 $=J6*Dx+K6$ となる。ただし、x は 6 から 189 までのそれぞれの行番号が入る。この場合、最初に、E6 のセルに $=J6*D6+K6$ を入れ、オートフィルを使って E189 までのセルに入力する。しかし、オートフィルを使うと、例えば、E7 のセルには、E6 と E7 の位置関係 = 一つ下 = 行番号が 1 増える関係から、数式内のセルの行番号も 1 増えてしまい、 $=J7*D7+K7$ になってしまう。この J7 とは、限界消費性向の推定値の標準誤差、K7 とは、定数項の推定値の標準誤差となり、全く意味をなさなくなる。一方、D7 はこの説明変数に対する予測値を E7 に入りたいので、これでいい。従って、J7, K7 と番号がずれないように、これらの変数については絶対参照を行う必要がある。つまり、E6 のセルには $=\$J\$6*D6+\$K\6 を入力しなければならない。その後、オートフィルを使う。

F6~F189: 残差を入力する。残差は、被説明変数の値 - その予測値であるから、 $=Cx-Ex$ となる。ただし、x には 6 から 189 までのそれぞれの行番号が入る。これを一度にやるためには F6 に $=C6-E6$ と入力した後、オートフィルを使う。

G6~G189: 標準化残差を入力する。標準化残差は、残差 / 残差標準偏差であるから、 $=Fx/SKS8$ である。ただし x には 6 から 189 までのそれぞれの行番号が入る。なぜ絶対参照が必要かは各々考えること。従って、まず G6 のセルに $=F6/SKS8$ を入力し、つぎにオートフィルで入力する。なお、Excel の分析ツールの標準化残差の計算は間違いであることがここからわかる。分析ツールを使用する場合は、この要領で正しい標準化残差を計算しよう。この中から、絶対値が 2.0 をこえるものを探すのは大変である。見つけやすくするには、後で述べる標準化残差のグラフをつくる方法と、絶対値が 2.0 を超える値を赤くいる図表する方法がある。後者は、まず標準化残差の領域をドラッグしておいて、ツールバーの書式を選び、でてきたメニューから「セル」を選ぶ。表示形式のタブを左クリックして種類の下の入力領域に、[赤]<=2]<0.000;[赤]>=2]>0.000;0.000 と入れ、OK をおす。

< 散布図 >




まず、ツールバーの中の「」をクリックし、グラフウィザードを起動し、グラフの種類から散布図を選ぶ。「次へ」をクリックしてする。次の画面では、「系列」のタグを選択する。まず、被説明変数と説明変数の散布図を造る。その画面の中から「追加」を選択し、「名前」の欄の右端の印を左クリックすると、入力エリアが現れる。そのとき、C5 のセル (Private Consumption と入っている) を左クリックし右上の x を左クリックする。つぎに、「X の値」の欄の右端の印を左クリックすると、入力エリアが現れるが、説明変数の範囲 (D6~D189) をドラッグして指定する。その後、右上の x を左クリックする。つぎは被説明変数を指定するが、「Y の値」の欄の右端の印を左クリックすると、入力エリアが現れるが、被説明変数の範囲 (C6~C189) をドラッグして指定する。そして完了を押す。これで散布図ができる。次は、説明変数と予測値の関係をこの散布図に入れるが、いまだ散布図のデータ系列のグラフを右クリックすると、メニューがでてくるが、この中から「近似曲線の追加」を選ぶ。そのまま OK をおすとできあがる。ただし、予測値を表す線 = 回帰直線の色が気に入らない場合は、それを右クリックしてでてきたメニューの中から近似曲線の書式設定を選び、好きな色を選べばよい。

< 説明変数 - 標準化残差の散布図 >

散布図と同様に作成する。ただし、近似曲線は必要ない。

< 標準化残差の時系列グラフ >



まず、ツールバーの中の「」をクリックし、グラフウィザードを起動し、グラフの種類から「折れ線」を選び、「次へ」をクリックする。「名前」の欄に散布図と同じ要領で「標準化残差」がはいっているセルを指定し、「値」の欄に、標準化残差の入っているセル (G6~G189) をドラッグして指定する。「項目軸ラベルに使用」の入力欄については、B6~B189 を指定する。そして、「完了」をおす。できたグラフの横軸 = 項目軸を右クリックして、「軸の書式設定」を選択する。新しいウィンドウのなかの「目盛」タグを左クリックする。その「目盛ラベルの間隔」に 7 を入力し、「配置」タブをクリックし、方向の角度を -90 度に設定する。さらに、「フォント」タブをクリックして、サイズの欄に 10 と入れる。そして、OK を左クリックする。