

## Excel による単純回帰分析

次のデータを分析する。

|   | A  | B  | C |
|---|----|----|---|
| 1 | Y  | X  |   |
| 2 | 1  | 1  |   |
| 3 | 4  | 2  |   |
| 4 | 9  | 4  |   |
| 5 | 15 | 6  |   |
| 6 | 20 | 10 |   |
| 7 |    |    |   |

エクセルにおいて、メニューバーの「データ」→「データ分析」→「回帰分析」を選択すると、次の画面が現れる。データ範囲、出力範囲を設定し、「OK」ボタンをクリックする。

参考: EXCEL2003 の場合は、メニューバーの「ツール」→「分析ツール」→「回帰分析」を選択すると、同様の画面が現れる。

回帰分析

入力元

入力 Y 範囲(Y):

入力 X 範囲(X):

☐ ラベル(L) ☐ 定数に 0 を使用(Z)

☐ 有意水準(O)  %

出力オプション

☐ 一覧の出力先(S):

☒ 新規ワークシート(P):

☐ 新規ブック(W)

残差

☐ 残差(R) ☐ 残差グラフの作成(D)

☐ 標準化された残差(T) ☐ 観測値グラフの作成(O)

正規確率

☐ 正規確率グラフの作成(N)

OK キャンセル ヘルプ(H)

図中の「入力Y範囲」には、被説明変数、「入力X範囲」には説明変数を示す範囲を入力する。ラベルには変数の名称をデータ範囲に含めて指定した際にチェックする。デ

一タ範囲にラベルを含めていないのにチェックすると、分析結果が違ってくるので注意する必要がある。出力先を指定したら、「OK」ボタンをクリックして、分析を行う。

実行すると、Excel は以下の分析結果を出力する。

|    | A      | B        | C        | D        | E        | F        | G        | H        | I        |
|----|--------|----------|----------|----------|----------|----------|----------|----------|----------|
| 1  | 概要     |          |          |          |          |          |          |          |          |
| 2  |        |          |          |          |          |          |          |          |          |
| 3  | 回帰統計   |          |          |          |          |          |          |          |          |
| 4  | 重相関 R  | 0.982995 |          |          |          |          |          |          |          |
| 5  | 重決定 R2 | 0.966279 |          |          |          |          |          |          |          |
| 6  | 補正 R2  | 0.955038 |          |          |          |          |          |          |          |
| 7  | 標準誤差   | 1.652019 |          |          |          |          |          |          |          |
| 8  | 観測数    | 5        |          |          |          |          |          |          |          |
| 9  |        |          |          |          |          |          |          |          |          |
| 10 | 分散分析表  |          |          |          |          |          |          |          |          |
| 11 |        | 自由度      | 変動       | 分散       | 割された分散   | 有意 F     |          |          |          |
| 12 | 回帰     | 1        | 234.6125 | 234.6125 | 85.96489 | 0.002655 |          |          |          |
| 13 | 残差     | 3        | 8.1875   | 2.729167 |          |          |          |          |          |
| 14 | 合計     | 4        | 242.8    |          |          |          |          |          |          |
| 15 |        |          |          |          |          |          |          |          |          |
| 16 |        | 係数       | 標準誤差     | t        | P-値      | 下限 95%   | 上限 95%   | 下限 95.0% | 上限 95.0% |
| 17 | 切片     | -0.04687 | 1.293734 | -0.03623 | 0.973373 | -4.16411 | 4.070363 | -4.16411 | 4.070363 |
| 18 | X 値 1  | 2.140625 | 0.230877 | 9.271725 | 0.002655 | 1.405872 | 2.875378 | 1.405872 | 2.875378 |
| 19 |        |          |          |          |          |          |          |          |          |
| 20 |        |          |          |          |          |          |          |          |          |
| 21 |        |          |          |          |          |          |          |          |          |

このままでは見にくいので、範囲を設定して、右クリックして「セルの書式設定」→「数値」→「少数点以下の桁数」を「4」に設定すると、つぎのようになる。

|    | A      | B       | C        | D        | E       | F       | G      | H        | I        |
|----|--------|---------|----------|----------|---------|---------|--------|----------|----------|
| 1  | 概要     |         |          |          |         |         |        |          |          |
| 2  |        |         |          |          |         |         |        |          |          |
| 3  | 回帰統計   |         |          |          |         |         |        |          |          |
| 4  | 重相関 R  | 0.9830  |          |          |         |         |        |          |          |
| 5  | 重決定 R2 | 0.9663  |          |          |         |         |        |          |          |
| 6  | 補正 R2  | 0.9550  |          |          |         |         |        |          |          |
| 7  | 標準誤差   | 1.6520  |          |          |         |         |        |          |          |
| 8  | 観測数    | 5.0000  |          |          |         |         |        |          |          |
| 9  |        |         |          |          |         |         |        |          |          |
| 10 | 分散分析表  |         |          |          |         |         |        |          |          |
| 11 |        | 自由度     | 変動       | 分散       | 割された分量  | 有意 F    |        |          |          |
| 12 | 回帰     | 1.0000  | 234.6125 | 234.6125 | 85.9649 | 0.0027  |        |          |          |
| 13 | 残差     | 3.0000  | 8.1875   | 2.7292   |         |         |        |          |          |
| 14 | 合計     | 4.0000  | 242.8000 |          |         |         |        |          |          |
| 15 |        |         |          |          |         |         |        |          |          |
| 16 |        | 係数      | 標準誤差     | t        | P-値     | 下限 95%  | 上限 95% | 下限 95.0% | 上限 95.0% |
| 17 | 切片     | -0.0469 | 1.2937   | -0.0362  | 0.9734  | -4.1641 | 4.0704 | -4.1641  | 4.0704   |
| 18 | X 値 1  | 2.1406  | 0.2309   | 9.2717   | 0.0027  | 1.4059  | 2.8754 | 1.4059   | 2.8754   |
| 19 |        |         |          |          |         |         |        |          |          |
| 20 |        |         |          |          |         |         |        |          |          |
| 21 |        |         |          |          |         |         |        |          |          |

「係数」のところには切片と傾きの値が示されている。推定された回帰式は、次のようになる。

$$Y = -0.0469 + 2.1406X, \quad R^2 = 0.9663$$

$$(-0.0362) \quad (9.2717)$$

ここで、括弧内の数値はt値である。エクセルのアウトプットにある「重決定 R2」は授業中で説明した「決定係数( $R^2$ )」に対応し、回帰モデルの当てはまりの良さを示す。決定係数は 0 から 1 の間の数値となり、1 に近いほど当てはまりが良いことを示している。上の例では、 $R^2 = 0.9663$  であり、当てはまりが良いことが分かる。

係数の欄の t 値, p 値は「 $H_0$ : 係数が 0 である」という帰無仮説を検定するために用いられる。検定結果から帰無仮説が棄却できない、すなわち統計的に 0 でないとはいえない(0 かもしれない)となると、Y と X の関係がないことになり、説明変数を含める意味がなくなってしまう。この検定は「有意性の検定」と呼ばれる。判断の仕方は以下の通りである。

(方法1) p 値 < 分析者が設定する有意水準 → 帰無仮説を棄却

(方法2) 検定統計量(t)の絶対値(|t|) > 分析者が設定した有意水準に対応するt分布の臨界値 → 帰無仮説を棄却

P値が求められている場合には、(方法1)で判断するほうが簡単である。上の例では、Xの p 値は 0.0027 と得られているので、 $0.0025 < 0.05$  より、有意水準5%で「Xの係数は0である」という帰無仮説は棄却され、Xは有意な説明変数であることがわかる。

#### 参考1:

t分布の確率値は、tdist 関数を用いて得ることができる。\$tdist(x,自由度,尾部)  
\$ ここで、x は、t 分布を計算する数値を指定する。自由度は、分布の自由度を整数で指定する。尾部は、片側分布を計算するか両側分布を計算するか、数値で指定する。尾部に 1 を指定すると片側分布の値が計算され、2 を指定すると両側分布の値が計算される。上の例では、x の値は 9.2717 であり、自由度は3、尾部は両側検定(2)であるので、 $\text{tdist}(9.2717, 3, 2) = 0.0027$  が得られる。

#### 参考2:

t 分布の臨界値は、tinv 関数を用いて得ることができる。\$tinv(有意水準、自由度)\$  
\$ ここで、有意水準には多くの場合、5%(0.05)、1%(0.01)といった値が用いられる。単純回帰モデルでは、自由度には、「n-2」の数が入る。ここで、n はサンプルサイズである。ここで、tinv 関数から得られる数値は両側検定のものである点に注意する必要がある。片側検定の値を求める場合は、引数に用いる有意水準を 2 倍する必要がある。例えば、5%片側の数字の場合は、0.1 となる。上の例では、X のt値は 9.271 であり、自由度は3である。 $\text{tinv}(0.05, 3) = 3.182$  より、 $|t\text{値}| > 3.182$  より、「x の係数が0である」という帰無仮説は棄却され、x は有意な説明変数であることがわかる。

---