

経済情報処理

GDP 産業計のデータに直線を当てはめる

データ

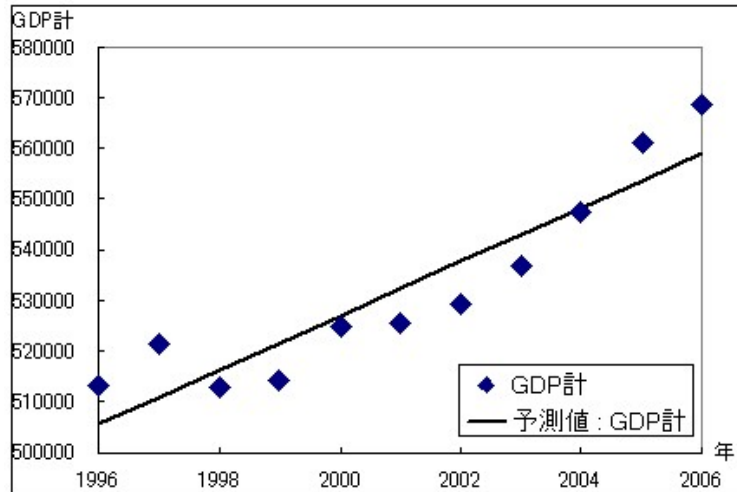
(1996, Y_0)、(1997, Y_1)、
(1998, Y_2)、 \dots 、(2006, Y_{10})

直線の方程式

$$Y = a + bX$$

$$\Rightarrow Y_i = a + bX_i$$

$$X_i = 1996, 1997,$$



• X_i と $T_i=1,2,3,\dots$ との違いは？

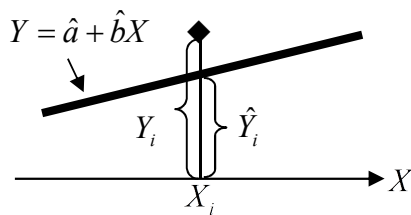
$$T_i = X_i - 1995 \rightarrow X_i = 1995 + T_i$$

$$Y_i = a + bX_i = a + b(1995 + T_i) = a + 1995b + bT_i$$

定数の推定

\hat{a}, \hat{b} : a, b の推定値

$$\hat{Y}_i = \hat{a} + \hat{b}X_i \quad (\text{注. } \hat{Y}_i \text{ は推定値、} Y_i \text{ は観測値})$$



最小二乗法

$\sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \hat{a} - \hat{b}X_i)^2$ の値が最も小さくなるような \hat{a}, \hat{b} を求める。

$$\hat{b} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}, \quad \hat{a} = \bar{Y} - \hat{b}\bar{X}$$

$e_i = Y_i - \hat{Y}_i$ を (回帰) 残差、 $\sum e_i^2$ を残差の二乗和という

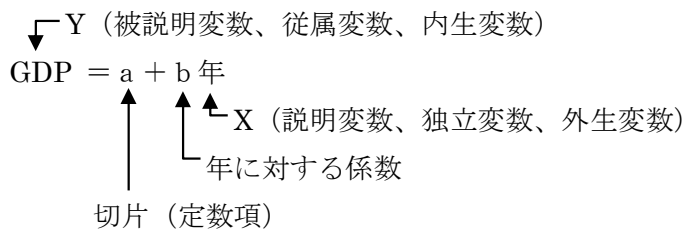
決定係数 R^2

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum e_i^2 \quad \text{ここで、} \bar{Y} = \frac{1}{n} \sum Y_i$$

全変動 説明できた 説明できなかった
 変動 変動

$$R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2} \quad 0 \leq R^2 \leq 1 \text{ (} R^2 \text{ は } Y_i \text{ と } X_i \text{ の相関係数の 2 乗)}$$

Excel による回帰分析



データは列方向に

(最初にラベルがあった方が分かりやすい)

Excel で回帰分析を行うには、ツール(T) ⇒ 分析ツール (2007 以降の場合は、データの
 リボン右端に 分析ツール がある) を指定し、開いたダイアログボックスで回帰分析を指
 定する。分析ツールが現れない場合は、ヒストグラムのところの説明した要領で登録する。

	A	B
1	年	GDP計
2	1996	513287.5
3	1997	521606.5
4	1998	512773.9
5	1999	514005.9
6	2000	524977.5
7	2001	525275.6
8	2002	529183.5
9	2003	536633.3
10	2004	547492.3
11	2005	561087.1
12	2006	568466.0

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(Q) 95 %

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(I) 観測値グラフの作成(I)

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)

概要								
回帰統計								
重相関 R	0.92195							
重決定 R2	0.849992							
補正 R2	0.833324							
標準誤差	7843.135							
観測数	11							
分散分析表								
	自由度	変動	分散	割された分散	有意 F			
回帰	1	3.14E+09	3.14E+09	50.99663	5.42E-05			
残差	9	5.54E+08	61514772					
合計	10	3.69E+09						
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	-1E+07	1496377	-6.78549	8.04E-05	-1.4E+07	-6768609	-1.4E+07	-6768609
年	5340.281	747.8136	7.141192	5.42E-05	3648.609	7031.953	3648.609	7031.953

一般の回帰分析

・ $Y_i = a + bX_i$ において、 X_i は 0,1,2,... といった連続数だけ？

・ Y_i の要因となるものならば

例 GDP=a+b 人口

この式の意味は、人口によって GDP が決定するという因果関係

厳密な定義

$$Y_i = a + bX_i + u_i$$

Y_i は基本的には X_i で決まるが、確率的に決まる u_i による誤差はある
 u_i を誤差項（攪乱項）

最小二乗法で望ましい結果が出る条件

- ・ X_i は確率変数でなく、確定した値
- ・ u_i は確率変数で期待値が 0
- ・ u_i の分散は i によらず一定
- ・ 異なる誤差項は無相関

回帰分析の結果の見方

・求められた $\hat{Y}_i = \hat{a} + \hat{b}X_i$ において、例えば、 $\hat{b} = 0$ という帰無仮説が棄却できれば

$H_0 : \hat{b} = c$ 、 $H_1 : \hat{b} \neq c$ という検定を行う。

$t = \frac{\hat{b} - c}{s.e.(\hat{b})}$ は自由度 $n-2$ の t -分布に従うことが分かっている。

$$\text{ここで、 } s.e.(\hat{b}) = \frac{s}{\sqrt{\sum (X_i - \bar{X})^2}}、s = \sqrt{\frac{\sum e_i^2}{n-2}} = \sqrt{\frac{\sum (Y_i - \hat{Y})^2}{n-2}} \text{ であるが、}$$

これは標準誤差の欄の値である。

従って、適当な c について t の値を計算し、設定した有意水準（多くの場合、1%、5%、10% が用いられる）の元で自由度 $n-2$ の t -分布表から棄却できるかどうかの両側検定を行えばよい（実は、Excel には t -分布の値を求める関数があり、それを用いればわざわざ t -分布表を用いる必要はないが、ここでは省略）。また、必要ならば対立仮説を $H_1 : \hat{b} < c$ などとして、片側検定を行ってもよい。

$H_0 : \hat{b} = 0$ という検定はよく行うので、 $t = \frac{\hat{b}}{s.e.(\hat{b})}$ の値は出力の t の部分に表示さ

れている。更に、その隣の P -値 というのは、自由度 $n-2$ の t -分布において $|t|$ よりも大きい（両側）部分の確率となっている（つまり、 P -値 が例えば 0.01 よりも小さければ、有意水準 1% の両側検定で棄却できることを意味する）。

\hat{a}, \hat{b} の求め方

$$\begin{aligned} \sum \frac{(Y_i - \hat{a} - \hat{b}X_i)^2}{\partial \hat{a}} &= \sum \frac{\partial \left((Y_i - \hat{b}X_i)^2 - 2(Y_i - \hat{b}X_i)\hat{a} + \hat{a}^2 \right)}{\partial \hat{a}} \\ &= \sum \left(\frac{\partial (Y_i - \hat{b}X_i)^2}{\partial \hat{a}} - \frac{\partial (2(Y_i - \hat{b}X_i)\hat{a})}{\partial \hat{a}} + \frac{\partial \hat{a}^2}{\partial \hat{a}} \right) \\ &= \sum (0 - 2(Y_i - \hat{b}X_i) + 2\hat{a}) \\ &= -2 \sum (Y_i - \hat{b}X_i) + 2 \sum \hat{a} = -2 \sum Y_i + 2\hat{b} \sum X_i + 2n\hat{a} = 0 \end{aligned}$$

$$\begin{aligned} \sum \left(\frac{\partial \left((Y_i - \hat{a} - \hat{b}X_i)^2 \right)}{\partial \hat{b}} \right) &= \sum \left(\frac{\partial \left((Y_i - \hat{a})^2 - 2(Y_i - \hat{a})\hat{b}X_i + \hat{b}^2 X_i^2 \right)}{\partial \hat{b}} \right) \\ &= \sum \left(\frac{\partial (Y_i - \hat{a})^2}{\partial \hat{b}} - \frac{\partial (2(Y_i - \hat{a})X_i\hat{b})}{\partial \hat{b}} + \frac{\partial X_i^2 \hat{b}^2}{\partial \hat{b}} \right) \\ &= \sum (0 - 2(Y_i - \hat{a})X_i + 2X_i^2 \hat{b}) \\ &= -2 \sum (Y_i X_i - \hat{a} X_i) + 2 \left(\sum X_i^2 \hat{b} \right) \\ &= -2 \sum Y_i X_i + 2 \left(\sum X_i \right) \hat{a} + 2 \left(\sum X_i^2 \right) \hat{b} = 0 \end{aligned}$$

結局、

$$\begin{cases} n\hat{a} + (\sum X_i)\hat{b} = \sum Y_i & (1) \\ (\sum X_i)\hat{a} + (\sum X_i^2)\hat{b} = \sum Y_i X_i & (2) \end{cases}$$

という連立方程式を解いて、 \hat{a} 及び \hat{b} を求めればよいということになる。

$$(1) \times \sum X_i - (2) \times n$$

$$\begin{array}{r} n(\sum X_i)\hat{a} + (\sum X_i)^2 \hat{b} = (\sum X_i)(\sum Y_i) \\ -) n(\sum X_i)\hat{a} + n(\sum X_i^2)\hat{b} = n\sum X_i Y_i \\ \hline ((\sum X_i)^2 - n\sum X_i^2)\hat{b} = (\sum X_i)(\sum Y_i) - n\sum X_i Y_i \end{array}$$

$$\begin{aligned}
& \text{ここで、 } \bar{X} = \frac{1}{n} \sum X_i, \bar{Y} = \frac{1}{n} \sum Y_i \text{ とすると、} \\
& n \sum (X_i - \bar{X})^2 = n \sum (X_i^2 - 2X_i\bar{X} + \bar{X}^2) = n \sum X_i^2 - 2n\bar{X} \sum X_i + n \sum \bar{X}^2 \\
& = n \sum X_i^2 - 2n^2\bar{X}^2 + n^2\bar{X}^2 = n \sum X_i^2 - n^2\bar{X}^2 = n \sum X_i^2 - n^2 \cdot \frac{1}{n^2} (\sum X_i)^2 \\
& = n \sum X_i^2 - (\sum X_i)^2 = -((\sum X_i)^2 - n \sum X_i^2) \\
& n \sum (X_i - \bar{X})(Y_i - \bar{Y}) = n \sum (X_i Y_i - X_i \bar{Y} - \bar{X} Y_i + \bar{X} \bar{Y}) \\
& \quad = n \sum X_i Y_i - n \bar{Y} \sum X_i - n \bar{X} \sum Y_i + n \sum \bar{X} \bar{Y} \\
& \quad = n \sum X_i Y_i - n^2 \bar{X} \bar{Y} - n^2 \bar{X} \bar{Y} + n^2 \bar{X} \bar{Y} \\
& \quad = n \sum X_i Y_i - n^2 \bar{X} \bar{Y} \\
& \quad = n \sum X_i Y_i - n^2 \frac{1}{n} (\sum X_i) \frac{1}{n} (\sum Y_i) \\
& \quad = n \sum X_i Y_i - (\sum X_i)(\sum Y_i) = -((\sum X_i)(\sum Y_i) - n \sum X_i Y_i)
\end{aligned}$$

$$\text{よって、 } \hat{b} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

また、(1)を変形すると、

$$\hat{a} = \frac{1}{n} \sum Y_i - \frac{1}{n} \hat{b} \sum X_i = \bar{Y} - \hat{b} \bar{X}$$

参考文献

縄田 和満 (1998) 「Excel による回帰分析入門」 朝倉書店