計量経済学 講義

第 8 回　　回帰分析 Part 2

2017 年 10 月 31 日（火）3 限

担当教員:　　唐渡　広志
研究室:　　経済学研究棟4階432号室
email:　　kkarato@eco.u-toyama.ac.jp
website:　　http://www3.u-toyama.ac.jp/kkarato/

講義の目的

- PCによる単純回帰分析の方法について学びます。
- べき乗関数や指数関数の回帰分析について学びます。

keywords: 分散分析表，自由度調整済み決定係数，分析ツール（Excel），両対数モデル，弾力性，半対数モデル，変化率

教科書: pp. 94 – 126（第3章）

【復習1】回帰分析

（単純）回帰分析とは

サンプルサイズ n のデータ $\{X_i\}, \{Y_i\}$ について回帰モデル

$$Y_i = \alpha + \beta X_i + u_i$$

の回帰パラメータ α, β の推定値を計算することである。

α, β の推定値の計算公式

α,β（真の値）と $\hat{\alpha}, \hat{\beta}$（推定値）は別のものなので区別をする

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} \qquad (3.1)$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X} \qquad (3.2)$$

推定回帰直線 $\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$ $\hat{Y}_i$ を理論値とよぶ。Y_i を実績値とよぶ。

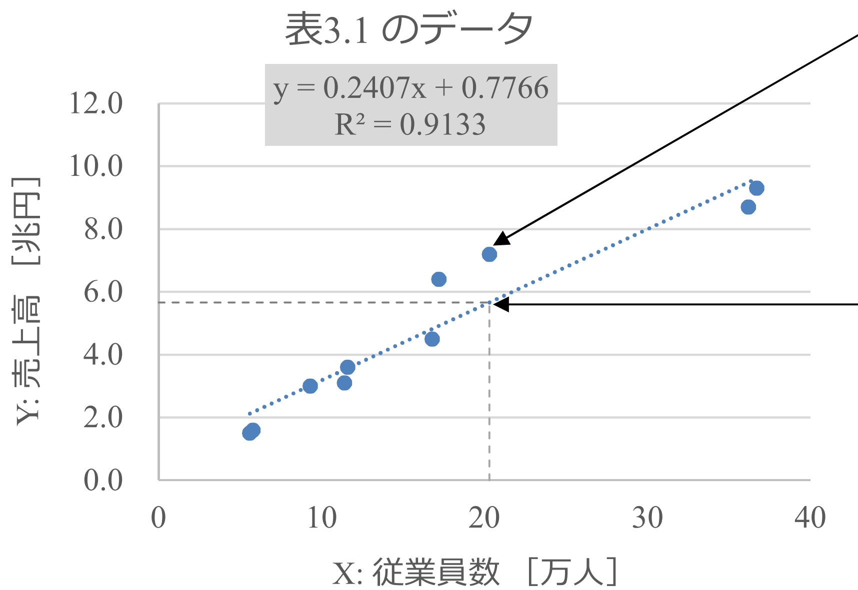
残差 $\hat{u}_i = Y_i - \hat{Y}_i \qquad (3.3)$ 残差 = 実績値 − 理論値

残差の特徴

$[ルール1]$. $\sum \hat{u}_i = 0 \qquad (3.5)$

$[ルール2]$. $\sum X_i \hat{u}_i = 0 \qquad (3.7)$

【復習2】回帰分析



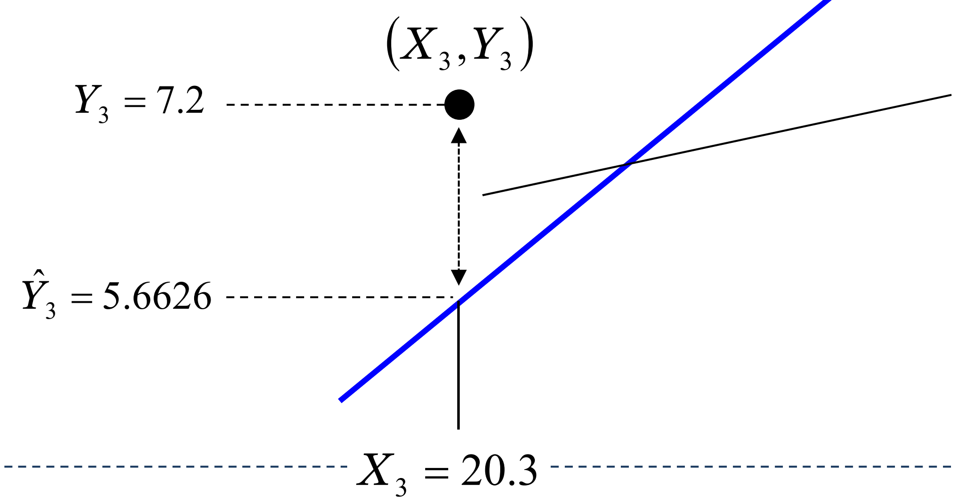
例. $i = 3$番目のデータ

$$(X_3, Y_3) = (20.3, 7.2)$$

Yの理論値(直線の高さ)

$$\hat{Y}_3 = 0.776 + 0.2407 \times 20.3 = 5.6626$$

$$\hat{Y}_i = 0.7766 + 0.2407X_i$$



残差

$$\hat{u}_3 = 7.2 - 5.6626 = 1.5374$$

【復習3】回帰分析

【残差2乗和】残差のばらつきを示す指標

$$\sum \hat{u}_i^2 = \underbrace{S_{yy}}_{\text{実績値のばらつき}} - \underbrace{S_{\hat{y}\hat{y}}}_{\text{理論値のばらつき}} \quad (3.15)\text{を参照}$$

【残差分散】残差のばらつきを示す指標

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} \qquad (3.17)$$

【回帰の標準誤差】

$$\hat{\sigma} = \sqrt{\frac{\sum \hat{u}_i^2}{n-2}}$$

【決定係数】推定回帰直線のあてはまりの良さを示す指標（アール・スクウェア）

$$R^2 = \frac{\sum\left(\hat{Y}_i - \overline{Y}\right)^2}{\sum\left(Y_i - \overline{Y}\right)^2} = \frac{S_{\hat{y}\hat{y}}}{S_{yy}} \qquad (3.9)$$

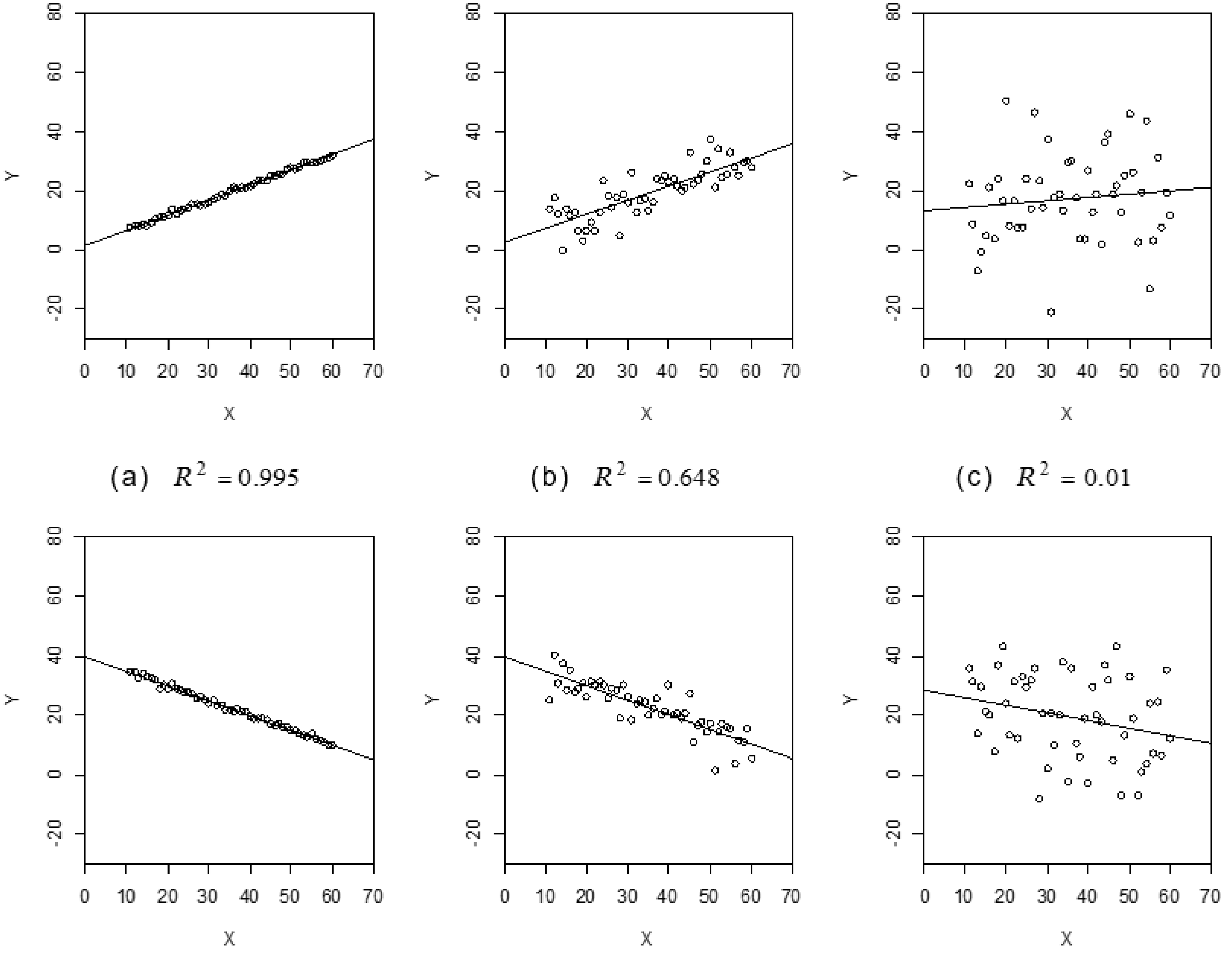
$$\text{または } R^2 = 1 - \frac{\sum \hat{u}_i^2}{S_{yy}} \qquad (3.16)$$

$$\text{決定係数の範囲}: 0 \le R^2 \le 1$$

【決定係数 =相関係数の2乗】

$$R^2 = \left(r_{xy}\right)^2 \qquad (3.10)$$

決定係数の例 p.103



(a) $R^2 = 0.995$　(b) $R^2 = 0.648$　(c) $R^2 = 0.01$

(d) $R^2 = 0.993$　(e) $R^2 = 0.719$　(f) $R^2 = 0.070$

図 3.3　決定係数

分散分析表 (ANOVA) (1) p.107

観測データ（実績値）における変動要因（ばらつき）を理論値の変動と残差の変動に分解すること。

$$(3.15) \rightarrow \underbrace{S_{\hat{y}\hat{y}}}_{\text{理論値のばらつき}} + \underbrace{\sum \hat{u}_i^2}_{\text{残差のばらつき}} = \underbrace{S_{yy}}_{\text{実績値のばらつき}}$$

変動/自由度

	自由度	変動（偏差２乗和）	分散	観測された分散比
回帰（理論値）	1（説明変数の数）	$S_{\hat{y}\hat{y}}$	$S_{\hat{y}\hat{y}}$	F 値: $\dfrac{S_{\hat{y}\hat{y}}}{\hat{\sigma}^2}$
残差	$n-2$	$\sum \hat{u}_i^2$	$\hat{\sigma}^2 = \dfrac{\sum \hat{u}_i^2}{n-2}$	-
合計（実績値）	$n-1$	S_{yy}	-	-

モデル全体の妥当性を測る指標
（F 統計量，後日詳しく勉強）

分散分析表 (ANOVA) (2) p.107

表3.1 のデータにおける 分散分析表

	従業員数[万人]	売上高[兆円]			
i	X_i	Y_i	残差^2	Yの偏差の2乗	理論値の偏差2乗
1	36.7	9.3	0.0961	19.4481	22.2780
2	36.2	8.7	0.6235	14.5161	21.1564
3	20.3	7.2	2.3635	5.3361	0.5969
4	17.2	6.4	2.2008	2.2801	0.0007
5	16.8	4.5	0.1025	0.1521	0.0049
6	11.6	3.6	0.0010	1.6641	1.7461
7	11.4	3.1	0.1768	3.2041	1.8756
8	9.3	3.0	0.0002	3.5721	3.5156
9	5.8	1.6	0.3279	10.8241	7.3843
10	5.6	1.5	0.3899	11.4921	7.6482
合計	170.9	48.9	6.2823	72.4890	66.2067
			$\sum \hat{u}_i^2$	S_{yy}	$S_{\hat{y}\hat{y}}$

	自由度	変動 (偏差2乗和)	分散	観測された分散比
回帰	1	66.207	66.207	84.309
残差	$n-2=8$	6.282	0.785	
合計	$n-1=9$	72.489		

自由度調整済み決定係数 (1)

自由度 $n-2$

$$R^2 = 1 - \frac{\sum \hat{u}_i^2}{S_{yy}} \qquad (3.16)$$

自由度 $n-1$

$\sum \hat{u}_i^2 / S_{yy}$ の分母・分子を $n-1$ で割ると

$$R^2 = 1 - \frac{\dfrac{\sum \hat{u}_i^2}{n-1}}{\dfrac{S_{yy}}{n-1}}$$

残差2乗和の自由度が誤っている。正しくは n-2

自由度調整済み決定係数
adjusted R-square
正しい自由度に調整した決定係数

$$adj.R^2 = 1 - \frac{\dfrac{\sum \hat{u}_i^2}{n-2}}{\dfrac{S_{yy}}{n-1}} = 1 - \frac{\hat{\sigma}^2}{s_y^2} \qquad (3.18)$$

自由度調整済み決定係数 (2)

表3.1 のデータの分散分析表と自由度調整済み決定係数

	自由度	変動 （偏差2乗和）	分散	観測された分散比
回帰	1	66.207	66.207	84.309
残差	$n-2=8$	6.282	0.785	
合計	$n-1=9$	72.489		

$\boxed{\sum \hat{u}_i^2}$ → 6.282

$\boxed{S_{yy}}$ → 72.489

$$\boxed{adj.R^2 = 1 - \frac{\dfrac{\sum \hat{u}_i^2}{n-2}}{\dfrac{S_{yy}}{n-1}} = 1 - \frac{\dfrac{6.282}{8}}{\dfrac{72.489}{9}} = 1 - \frac{0.785}{8.054} = 0.9025}$$

$$\boxed{R^2 = 1 - \frac{\sum \hat{u}_i^2}{S_{yy}} = 1 - \frac{6.282}{72.489} = 1 - 0.0867 = 0.9133}$$

自由度調整済み決定係数 (3)

R^2 と $adj.R^2$ の比較

$$adj.R^2 = 1 - \frac{\dfrac{\sum \hat{u}_i^2}{n-2}}{\dfrac{S_{yy}}{n-1}}$$

$$= 1 - \frac{n-1}{n-2} \cdot \frac{\sum \hat{u}_i^2}{S_{yy}}$$

$$= 1 - \frac{n-1}{n-2} \cdot \left(1 - R^2\right)$$

$$R^2 = 1 - \frac{\sum \hat{u}_i^2}{S_{yy}}$$

$$\Leftrightarrow \quad adj.R^2 - R^2 = \frac{1}{n-2}\left(R^2 - 1\right) \leq 0$$

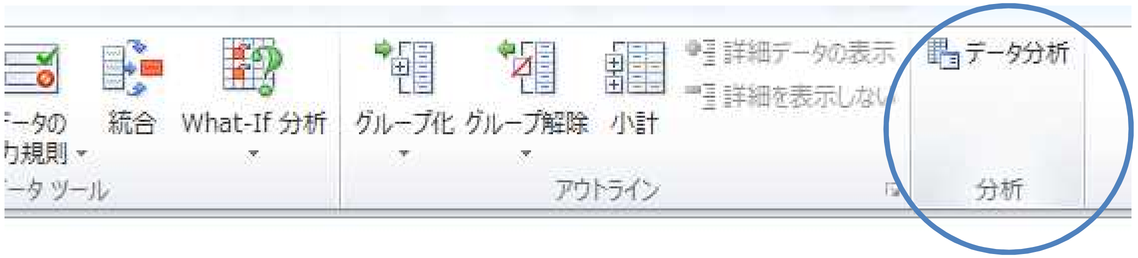
$$\Leftrightarrow \quad adj.R^2 \leq R^2$$

決定係数の範囲 : $0 \leq R^2 \leq 1$

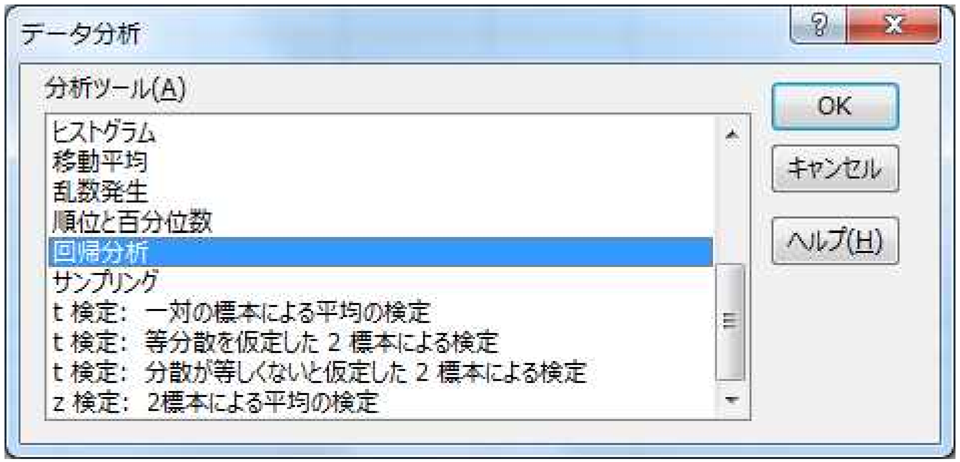
自由度調整済み決定係数は決定係数よりも小さめの値になる

Excel「分析ツール」(1)

1. 「データ」タブ→「データ分析」



2. 「データ分析」ダイアログの中から「回帰分析」を選択



Excel「分析ツール」(2)

3.「回帰分析」ダイアログにおいてYのデータとXのデータを指定

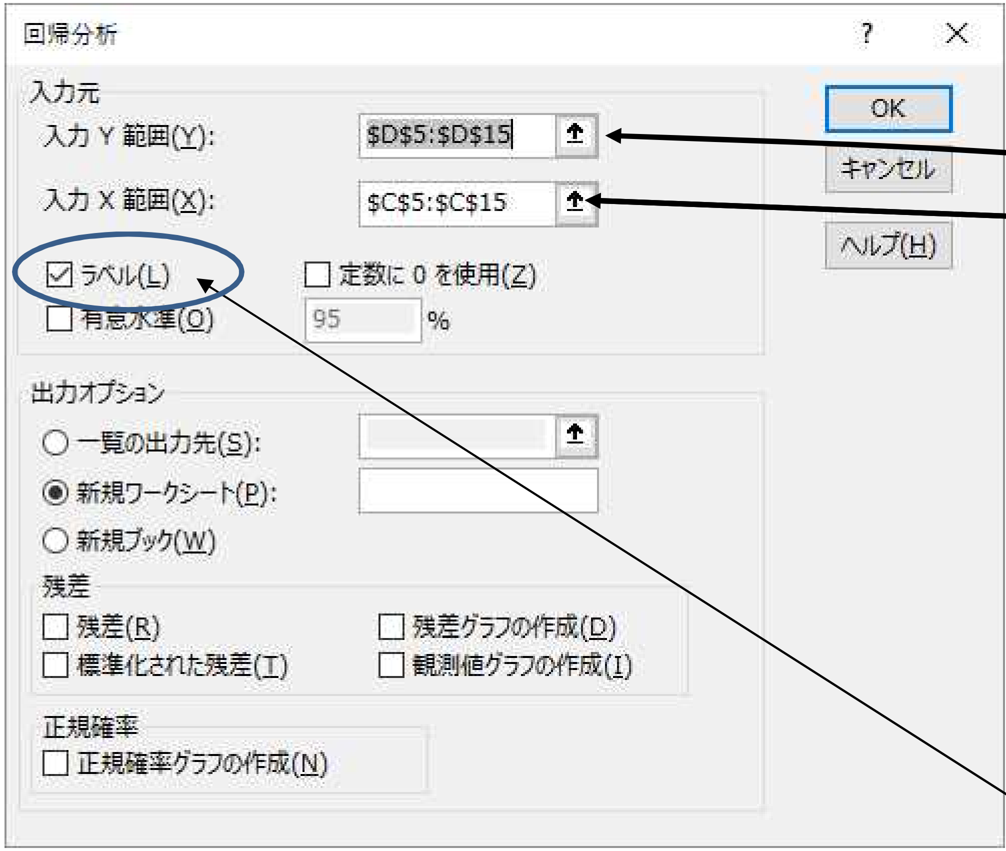


表3.1のデータ

X_i	Y_i
36.7	9.3
36.2	8.7
20.3	7.2
17.2	6.4
16.8	4.5
11.6	3.6
11.4	3.1
9.3	3.0
5.8	1.6
5.6	1.5

ラベル（変数名X,Y）を含める場合は，チェック

Excel「分析ツール」(3)

4. 「回帰分析」の出力結果

回帰統計	意味
重相関 R	$\{Y_i\}$ と $\{\hat{Y}_i\}$ または $\{X_i\}$ との相関係数 $r_{y\hat{y}}$ または r_{xy}
重決定 R2	決定係数 R^2
補正 R2	自由度調整済み決定係数 $adj.R^2$
標準誤差	回帰の標準誤差 $\hat{\sigma}$
観測値	サンプルサイズ n

	A	B	C	D	E	F	G	H	I
1	概要								
2									
3	回帰統計								
4	重相関 R	0.955685							
5	重決定 R2	0.913335							
6	補正 R2	0.902501							
7	標準誤差	0.886163							
8	観測数	10							
9									
10	分散分析表								
11		自由度	変動	分散	測された分散	有意 F			
12	回帰	1	66.20672	66.20672	84.3091	1.6E-05			
13	残差	8	6.282284	0.785286					
14	合計	9	72.489						
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	0.776586	0.528413	1.469655	0.179852	-0.44194	1.995109	-0.44194	1.995109
18	Xi	0.240691	0.026213	9.181999	1.6E-05	0.180243	0.301139	0.180243	0.301139
19									

$\hat{\alpha}$

$\hat{\beta}$

分散分析表

後日勉強

推定結果のシンプルな報告方法

$$\hat{Y}_i = \boxed{\mathbf{0.78}} + \boxed{\mathbf{0.24}}\, X_i,$$

$$R^2 = \boxed{\mathbf{0.913}},\ adj.R^2 = \boxed{\mathbf{0.903}},\ \hat{\sigma} = \boxed{\mathbf{0.89}}$$

練習問題 (1)

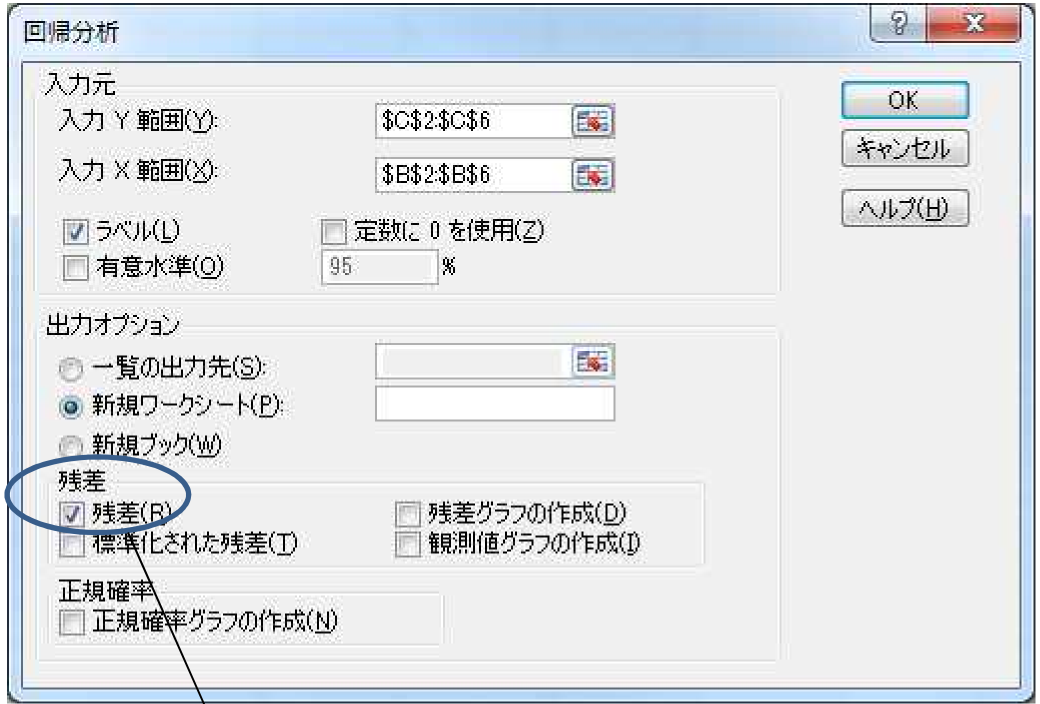
- p.142, 表3.16 製造業（企業規模：10～99人）の生産労働者（高校卒，男性）の年齢Xと賃金Y のデータについて，「分析ツール」の「回帰分析」を利用して推定を行い以下の空欄を埋めなさい。

$$\hat{Y}_i = \boxed{} + \boxed{} X_i,$$

$$R^2 = \boxed{}, adj.R^2 = \boxed{}, \hat{\sigma} = \boxed{}$$

Excel「分析ツール」(4)

■理論値と残差の出力



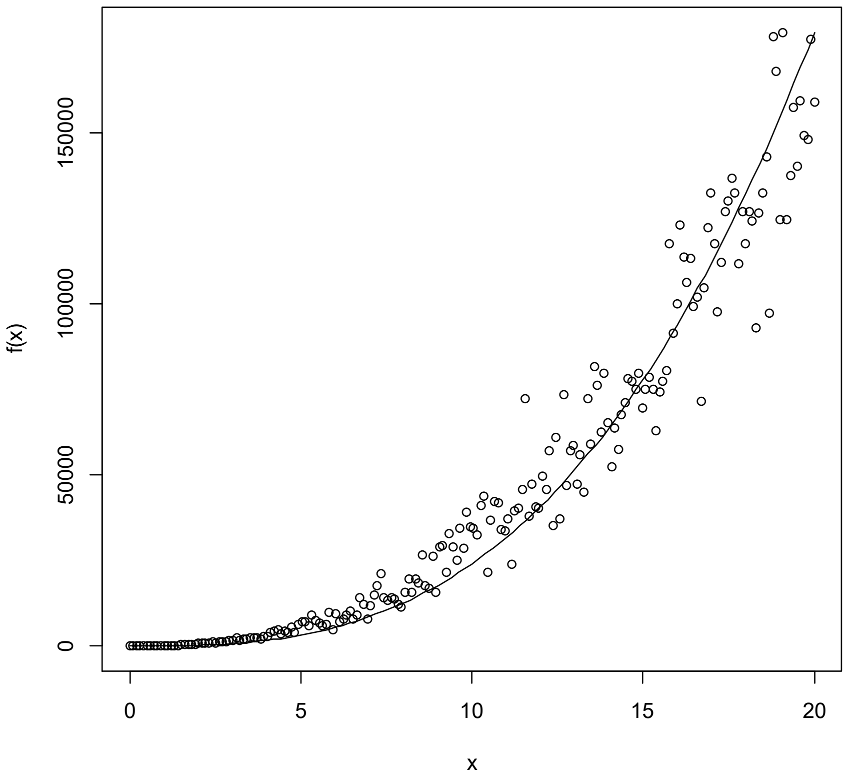
「残差」にチェック

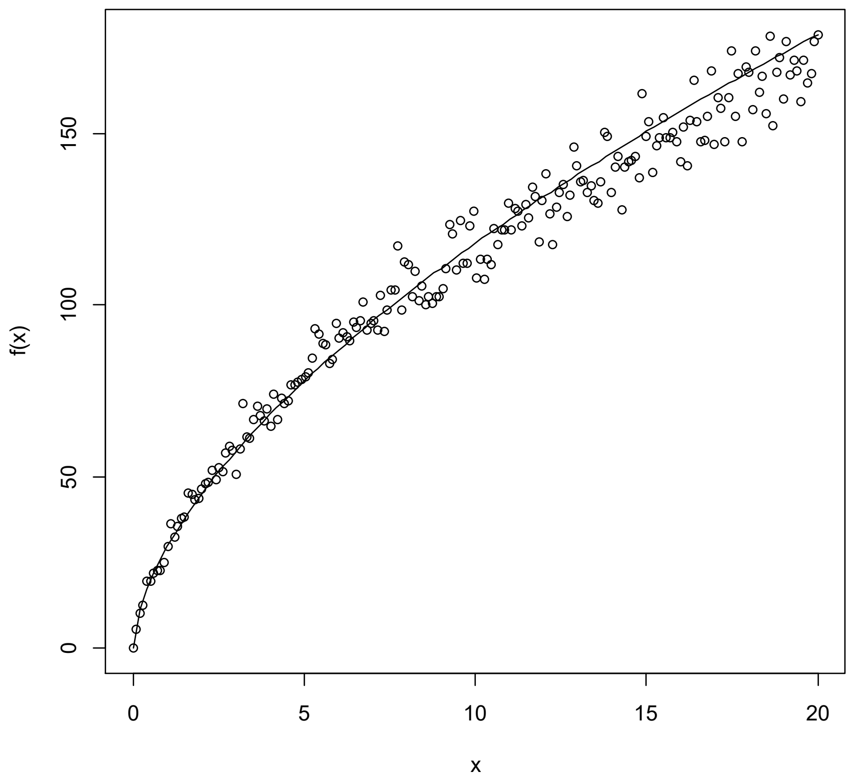
理論値 $\hat{Y}_i$ 残差 $\hat{u}_i$

残差出力

観測値	予測値: Yi	残差
1	9.609956	−0.30996
2	9.489611	−0.78961
3	5.662619	1.537381
4	4.916476	1.483524
5	4.8202	−0.3202
6	3.568605	0.031395
7	3.520466	−0.42047
8	3.015015	−0.01501
9	2.172595	−0.5726
10	2.124457	−0.62446

非線形式の回帰分析





$y = ax^b$ や $y = ae^{cx}$ のような式の

a, b, c を求めるにはどうしたらよいか？

指数関数

$$y = c^x$$

（xを指数とよぶ。$c > 1$ は定数であり，指数の底 [base] とよばれる）

$c = 2$ の例.

$$y = 2^4 = 16$$

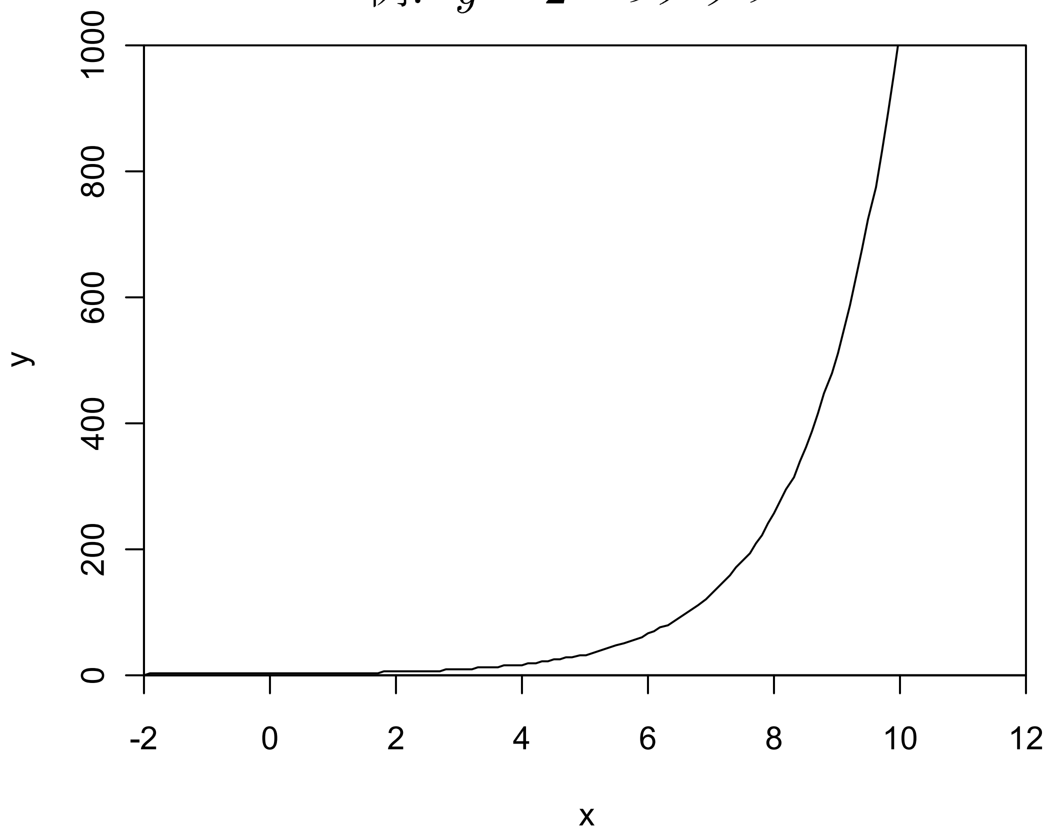
$$y = 2^{0.5} = 1.414\cdots \ (= \sqrt{2})$$

$$y = 2^0 = 1$$

$$y = 2^{-1} = \frac{1}{2}$$

$$y = 2^{-4} = \frac{1}{16}$$

例. $y = 2^x$ のグラフ



【復習】特殊な底

$y = e^x$ 　($y = \exp(x)$ と書く場合もある)

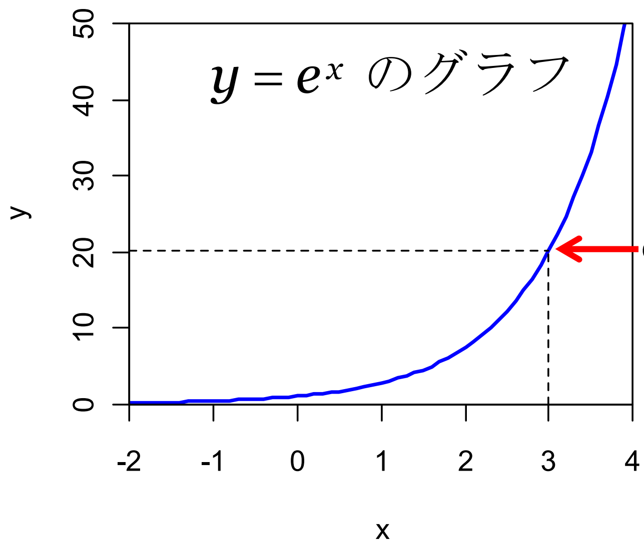
底が e の指数関数（**自然指数関数**）

十分に大きな m に対して

$$e = \left(1 + \frac{1}{m}\right)^m = 2.71828\cdots$$

となる e を自然対数の底（またはネイピア数）とよぶ.

m	e
1	2
10	2.59374246
100	2.70481383
1,000	2.71692393
10,000	2.71814593
100,000	2.71826824
1,000,000	2.71828047
10,000,000	2.71828169
100,000,000	2.71828179
1,000,000,000	2.71828203



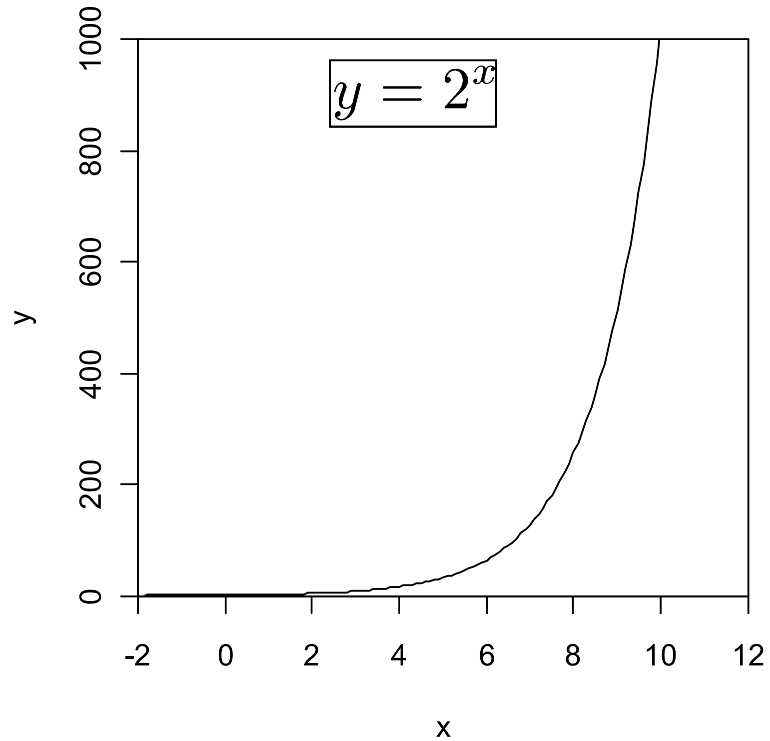
$\exp(3) = 20.08554$

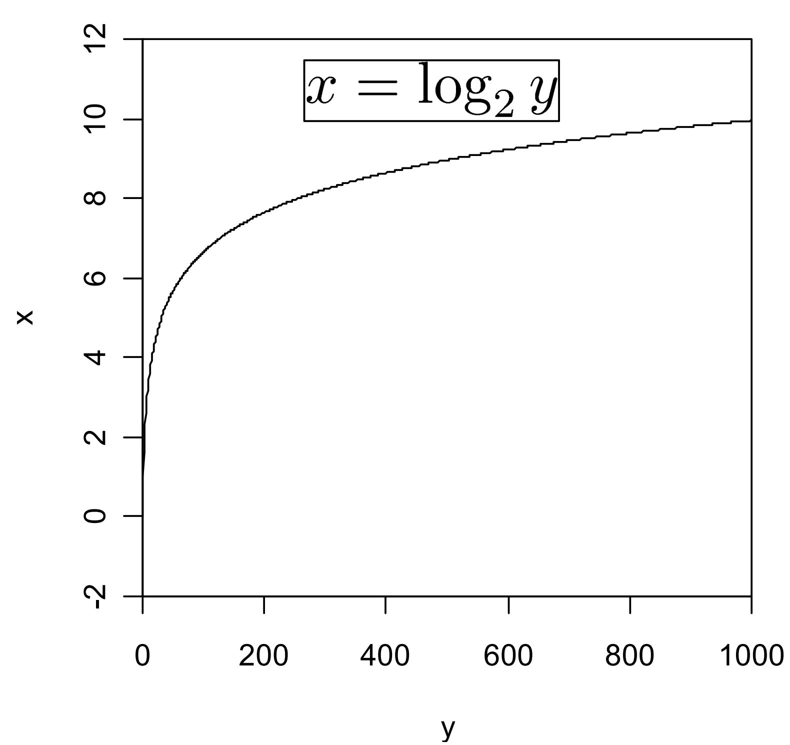
【復習】指数関数の逆関数：対数

$y = c^x$ を $x = \cdots$ の形にするには？

$$\boxed{\log_c y = \log_c(c^x) \leftrightarrow \log_c y = x}$$

- y が c を底とする x の指数関数 $y=c^x$ であるとき，x の値は c を底とする y の対数 $\log_c y = x$ に等しい。
- 指数関数の逆関数をとるとき，指数の底 c は y の<u>対数の底</u>になる．



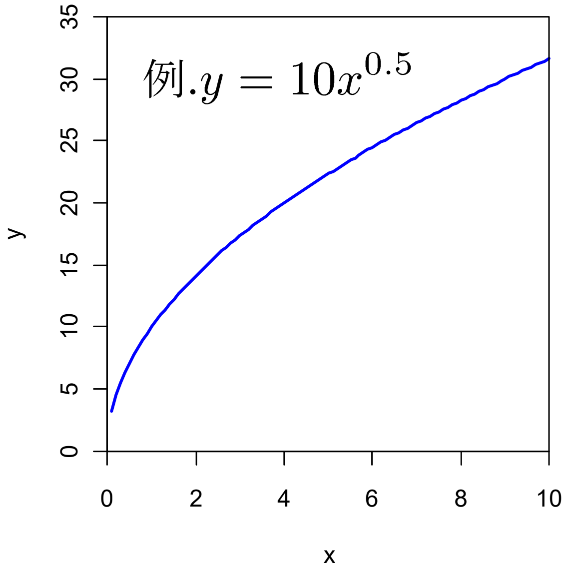


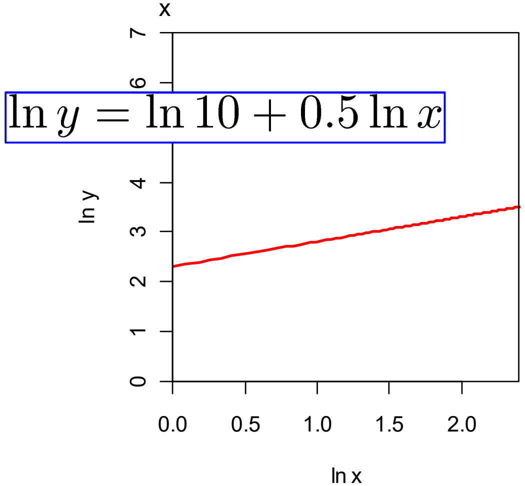
ベキ関数: $y = ax^b$ の対数線型化 (pp.115-119)

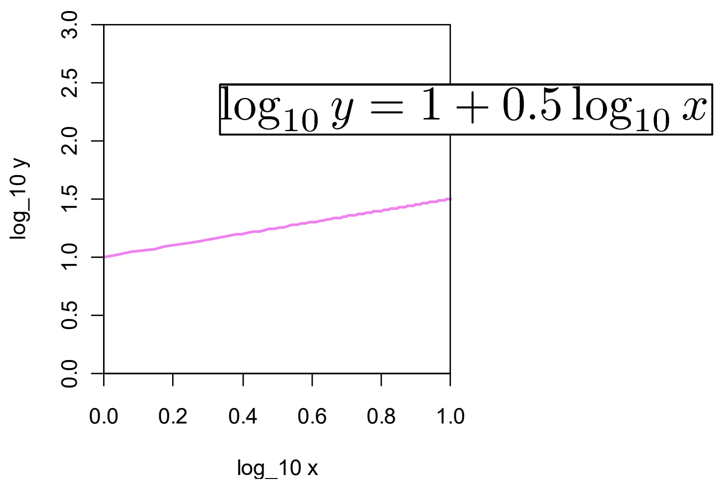
$y = ax^b \quad \rightarrow \quad \ln y = \ln a + b \ln x$

または $\quad \rightarrow \quad \log_{10} y = \log_{10} a + b \log_{10} x$

x	y	$\ln x$	$\ln y$	$\log_{10} x$	$\log_{10} y$
1	10	0	2.303	0	1
2	14.142	0.693	2.649	0.301	1.151
3	17.321	1.099	2.852	0.477	1.239
4	20	1.386	2.996	0.602	1.301
5	22.361	1.609	3.107	0.699	1.349
6	24.495	1.792	3.198	0.778	1.389
7	26.458	1.946	3.276	0.845	1.423
8	28.284	2.079	3.342	0.903	1.452
9	30	2.197	3.401	0.954	1.477
10	31.623	2.303	3.454	1	1.5







練習問題 (2)

[1] 以下の x の指数関数について底 をcとして両辺の対数をと りなさい。

[1.1] $y = c^{2x}$

[1.2] $y = 10c^{-x}$

[2] 以下の x の指数関数について底 をe(ネイピア数) として 両辺の対数をとりなさ い。

[2.1] $y = 2e^{x}$

[2.2] $y = e^{\alpha+\beta x}e^{-\gamma x}$

[3] 以下の ベキ関数について両辺 の対数をとりなさい。

[3.1] $y = 10x^{0.2}$

[3.2] $y = e^{2}x^{-2}$

両対数モデル (1)

$$y = ax^b \quad \rightarrow \quad \ln y = \ln a + b \ln x$$

変数の置き換え: $\ln y \rightarrow \boxed{Y_i}$, $\ln a \rightarrow \alpha$, $b \rightarrow \beta$, $\ln x \rightarrow \boxed{X_i}$

- 両辺の変数に対数が使用されているので両対数（ダブルログ）モデルとよぶ。
- 両対数モデルは $\ln y$ と $\ln x$ の直線的な関係をみている。つまり，$\ln x$ が1単位増えると，$\ln y$ は b だけ変化する。

対数変換した上で $Y_i = \alpha + \beta X_i + u_i$ を推定

a の求め方

$$\alpha = \ln a \rightarrow \hat{a} = \exp(\hat{\alpha})$$

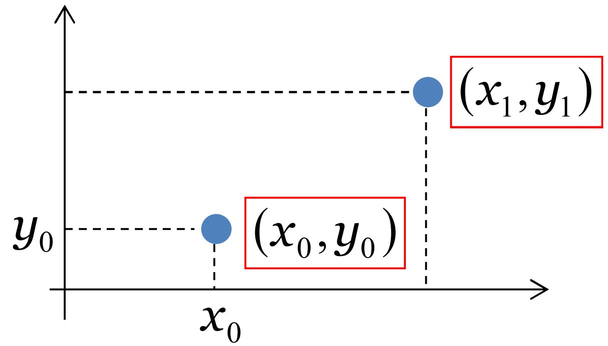
表3.7 (p.117) を参照

両対数モデルの傾き b (β) は y の x に関する弾力性を示している

両対数モデル (2)：弾力性

$$y = ax^b \quad \rightarrow \quad \ln y = \ln a + b \ln x$$

b の意味： $\ln x$ が 1 単位増えるときの $\ln y$ の変化量？



$$\ln y_1 = \ln a + b \ln x_1$$
$$\ln y_0 = \ln a + b \ln x_0$$

引き算

$$\ln y_1 - \ln y_0 = b(\ln x_1 - \ln x_0)$$

$$\boxed{b = \frac{\ln y_1 - \ln y_0}{\ln x_1 - \ln x_0} = \frac{y\text{ の変化率(の近似)}}{x\text{ の変化率(の近似)}}}$$

【復習】対数差分は変化率の近似

【重要】x の変化率に対する y の変化率の比 (=b) のことを「 y の x （に関する）**弾力性**」とよぶ。

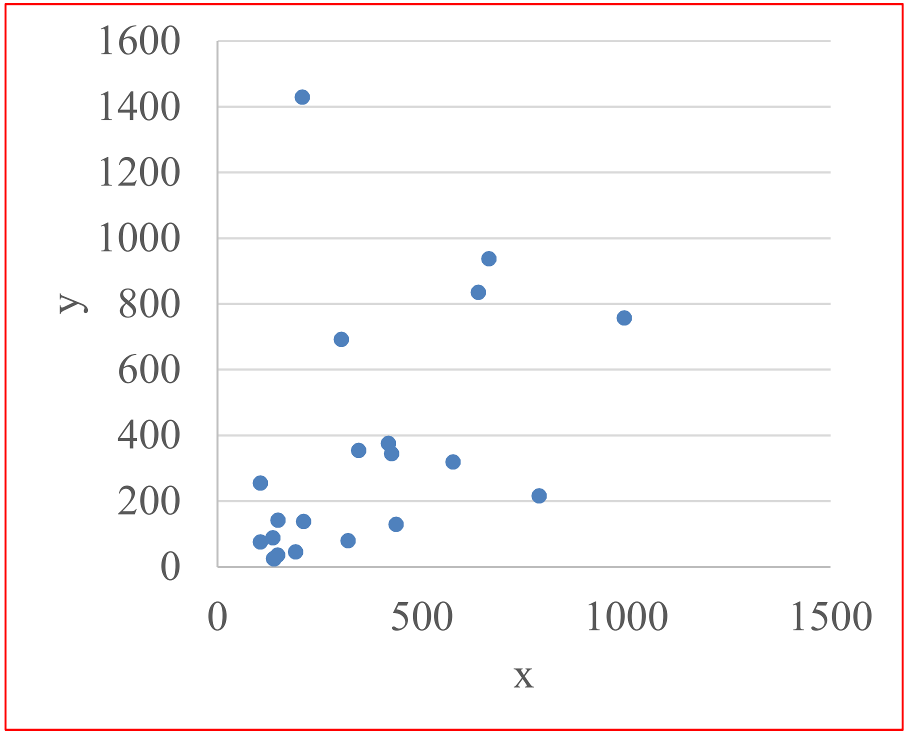
意味→弾力性 b は **x が1%変化したときの y の変化率を示している**。$b > 1$ ならば弾力的，$b < 1$ ならば非弾力的。両対数モデルの傾き b は「弾力性」を求めるのに便利な形をしている。

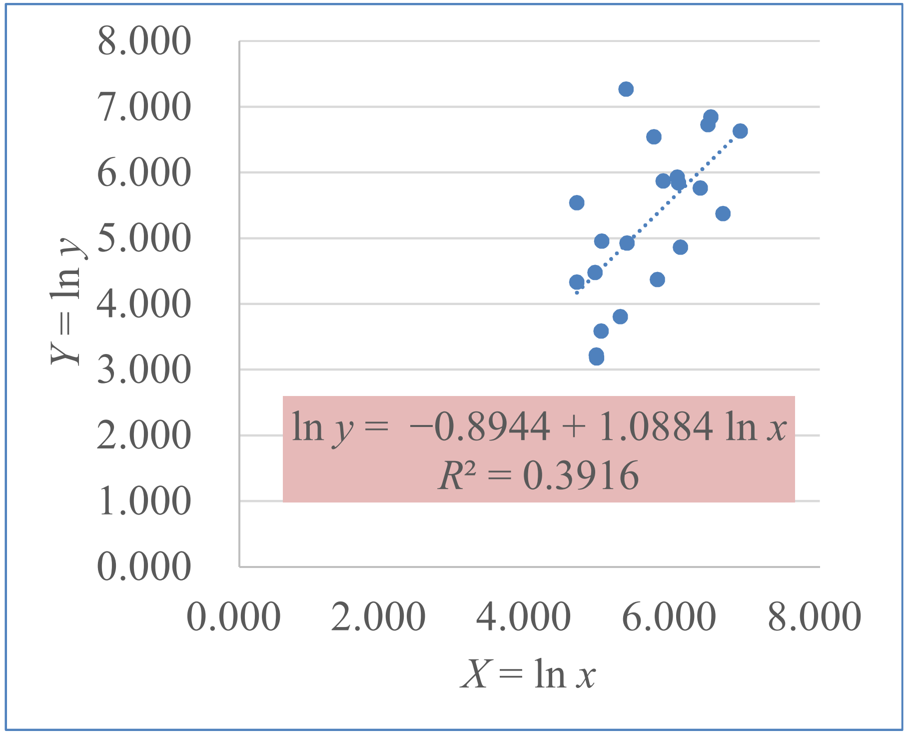
例. 需要の価格弾力性：（需要関数において数量の変化率/価格の変化率）

練習問題 (3)

表3.7 (p.117)のデータについて，従業員数 x，売上高 y を用いてベキ乗関数 $y = ax^b$ を推定しなさい。

$$y = ax^b \to \ln y = \ln a + b \ln x$$





$$\hat{a} = \exp(\hat{\alpha}) = \exp(-0.8944) = 0.409 \longrightarrow \hat{y}_i = 0.409 x_i^{1.088}$$

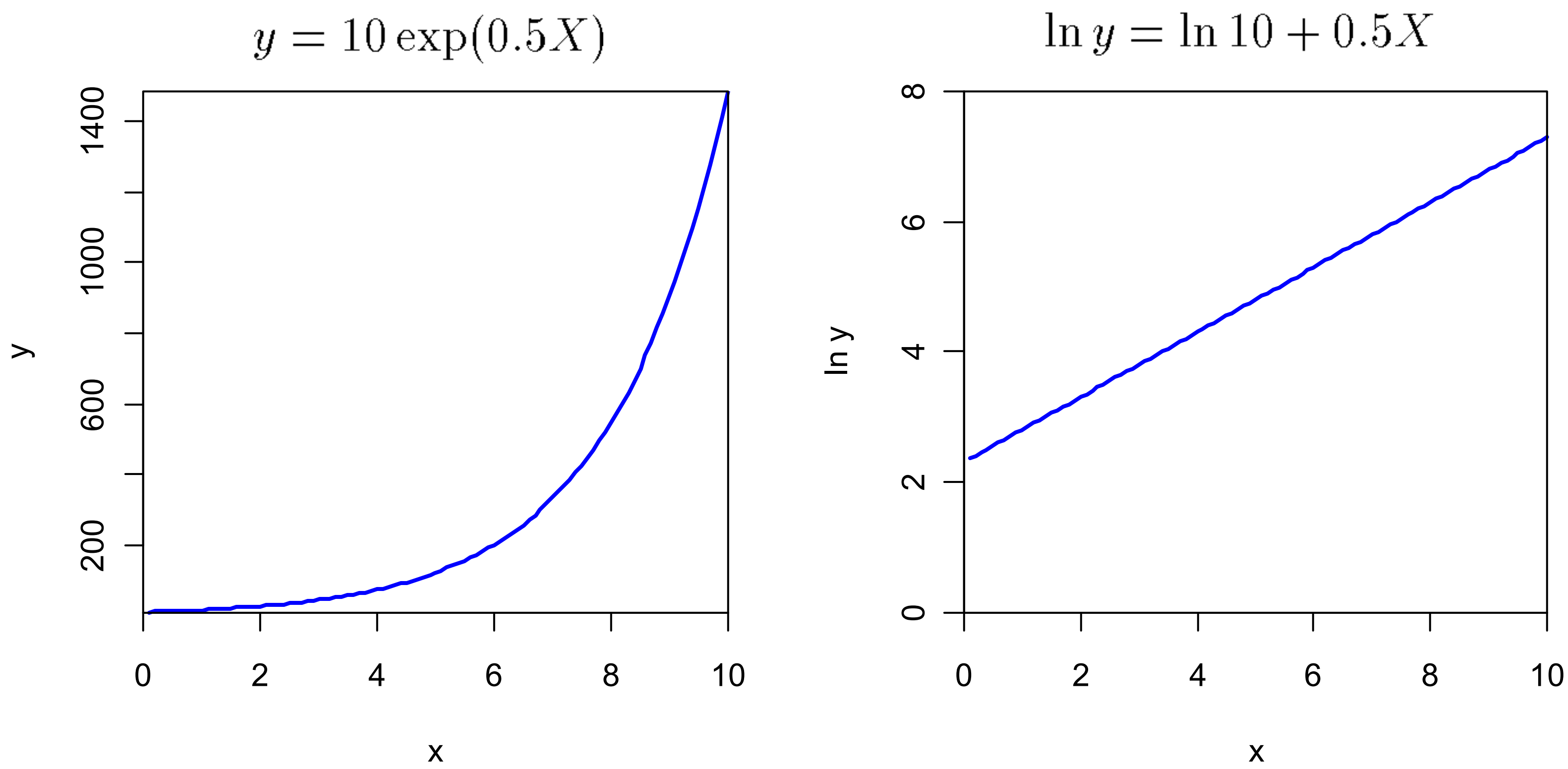
$$\hat{b} = 1.0884$$

意味：x（従業員数）が1%増えると，売上高 y は1.0884%増える。

自然指数関数: $y = ae^{cX}$ の対数線型化 (pp.119-122)

$$y = ae^{cX} = a\exp(cX) \quad \rightarrow \quad \ln y = \ln a + cX$$

X が 1 単位増えると，$\ln y$ は0.5増加する



半対数モデル (1)

$$y = ae^{cX} = \quad \rightarrow \quad \underbrace{\ln y}_{\boxed{Y_i}} = \underbrace{\ln a}_{\alpha} + \underbrace{c}_{\beta}\underbrace{X}_{\boxed{X_i}}$$

変数の置き換え

x はそのまま使える

- 左辺にだけ対数が使用されているので半対数（セミログ）モデルとよぶ。
- 半対数モデルは $\ln y$ と x の直線的な関係をみている。x が1単位増えると，$\ln y$ は c だけ変化する。

表3.8 (p.120) を参照

半対数モデルの傾き c は y の X に関する変化率を示している

半対数モデル (2)：変化率の測定

$y = ae^{cX} = \quad \rightarrow \quad \ln y = \ln a + cX$

c の意味： X が 1 単位増えるときの $\ln y$ の変化量？

$$\ln y_1 = \ln a + cX_1$$
$$\ln y_0 = \ln a + cX_0$$

引き算

$$\boxed{\ln y_1 - \ln y_0 = c(X_1 - X_0)} \longrightarrow \boxed{X_1 - X_0 = 1 \text{ ならば } c = \ln y_1 - \ln y_0}$$

【復習】対数差分は変化率の近似

【重要】半対数モデルの回帰パラメータ c は X が 1 単位増えるときの，変化率の近似値（y の対数差分）になっている。

$X_1 - X_0 = 1$ ならば

$$c = \ln y_1 - \ln y_0 = \ln \frac{y_1}{y_0} \Leftrightarrow \exp(c) = \frac{y_1}{y_0}$$
$$\Leftrightarrow \exp(c) - 1 = \frac{y_1}{y_0} - 1$$
$$\Leftrightarrow \frac{y_1 - y_0}{y_0} = \exp(c) - 1$$

正確な変化率

練習問題 (4)

表3.8 (p.120)のデータについて，年次 X，人口 y を用いて指数関数 $y = ae^{cX}$ を推定しなさい。

$$\boxed{\ln y_i = \ln a + cX_i + u_i} \xrightarrow{\text{推定結果}} \boxed{\hat{Y}_i = -27.612 + 0.021X_i}$$

$\hat{\alpha}$($=\ln a$の推定値)からaの推定値を計算する

$$\hat{a} = \exp(\hat{\alpha}) = \exp(-27.612) = 1.0195\text{E-}12$$

$0.0000000000010195 = 10^{-12} \times 1.0195$

E-12→小数点12桁目以降にゼロ以外の数値が初めて出る小さい値

以上より指数関数は

$$\boxed{\hat{y}_i = 10^{-12} \times 1.0195 e^{0.021X_i}}$$

パラメータ c の意味

$$\ln y_i - \ln y_{i-1} = c(X_i - X_{i-1})$$

$$\leftrightarrow \ln \frac{y_i}{y_{i-1}} = c(X_i - X_{i-1})$$

$$\leftrightarrow \frac{y_i}{y_{i-1}} = e^{c(X_i - X_{i-1})}$$

$$\leftrightarrow \frac{y_i - y_{i-1}}{y_{i-1}} = e^{c(X_i - X_{i-1})} - 1$$

$\hat{c} = 0.021$,

$X_i - X_{i-1} = 5$ [年]より

$$\boxed{\frac{y_i - y_{i-1}}{y_{i-1}} = e^{0.021\times 5} - 1 = 0.1107}$$

5年あたりの人口成長率は11.07%

1年あたりでは $e^{0.021}-1 = 0.0212$ (2.12%)

練習問題 (5)

[1]. y: 一人あたりGDP, x: 一人当たり資本装備率とする. 1980-2008年のデータ（サンプルサイズは29）を利用してべき関数 $y = ax^b$, を両対数モデルで回帰分析を行ったところ次の結果が得られた。一人あたり資本装備率が 1% 増えると, 一人当たりGDPは何%増えるか?

また, y の x に関する弾力性の値を答えなさい。

$$\ln\hat{y}_i = -6.47 + 0.69\ln x_i, \quad R^2 = 0.965, \quad adj.R^2 = 0.947$$

[2]. 関西地方の2012年7月平日(n = 23日)の最高気温（℃）Xと電力消費量（万kwh）yを用いて指数関数 $y = ae^{cX}$を推定する。半対数モデルで回帰分析を行ったところ次の結果が得られた。気温が1度上昇すると, 電力消費量は何%変化すると言えるか？

$$\ln\hat{y}_i = 6.91 + 0.026X_i, \quad R^2 = 0.652, \quad adj.R^2 = 0.635$$

両対数，半対数モデルの傾きについてのまとめ

ベキ関数から両対数モデルへ

$$y = ax^b \quad \rightarrow \quad \ln y = \ln a + b \ln x$$

b の意味：<u>x が 1 % 増えるときの</u> y の変化率 %

例. $\ln y = 2.42 + 0.75 \ln x$

→ x が 1%増えると，y は 0.75% 増える。

底を e とする指数関数から半対数モデルへ

単位に注意が必要

$$y = ae^{cX} = \quad \rightarrow \quad \ln y = \ln a + cX$$

c の意味：x が <u>1 単位増えるときの</u> y の変化率 %

例. $\ln y = 0.921 + 0.024\, X$

→ X が 1 単位増えると，y は 2.4% 増える。

まとめ

- Excel で（単純）回帰分析を行う場合，関数を利用する方法，散布図に直接書き込む方法，「分析ツール」を利用する方法がある。
- 変数に関して非線型であっても，変数変換によって標準モデルに直すことができれば，最小2乗法が適用可能である。
- 両対数モデルの傾きから変数間の弾力性を求めることができる。
  - 弾力性とは X の変化率に対する Y の変化率の比
- 自然指数関数を対数化した半対数モデルの傾きは成長率（変化率）に対応している。
  - X が1単位変化したときの Y の変化率