計量経済学 講義

第 2 回 経済データの扱い方 Part 1

2017 年 10 月 6 日（金）1 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

■ 経済データを分類し，それぞれの特徴に応じた扱い方について学びます。

Keywords : 質的データ，量的データ，クロスセクションデータ，時系列データ，クロス集計表，変化率，対数変換，対数差分

教科書：pp. 1—9（第1章）

データの収集

（例）既婚女性の就業状況に関するアンケート調査：
調査時点　20XX年〇月

- a. 年齢
- b. 現在仕事をしているか？
  （1.はい，2.いいえ）
- c. 現在の仕事に満足しているか？
  （1. 満足している，2. やや満足している，
  3. どちらともいえない，4. やや不満足，5. 不満足）
- d. 現在の職種
  （1.専門・技術職　2.管理職　3.事務職　4.販売職・・・）
- e. 18歳未満の子供の数
- f. 夫の年収［万円］
- g. 婚姻した時期（西暦年）

・
・
・
・
・
・

データの集計（表1.1, p.2）

	a	b	c	d	e	f	g
個体番号	年齢	就業状況	仕事の満足度	職業	18歳未満の子供の数	夫の年収[万円]	婚姻時期[年]
1	39	1	2	1	0	530	2004
2	30	1	3	5	2	460	2007
3	52	2	-	-	0	880	1987
4	36	1	2	5	2	750	2002
5	44	1	3	2	1	480	1997
6	66	2	-	-	0	230	1971
7	27	1	2	3	0	390	2008
8	29	2	-	-	3	540	1999
9	54	1	1	4	0	650	1986
10	25	1	1	5	1	340	2009

質的データ / カテゴリカル・データ

［質的 = Qualitative］数や量では測れないデータ

物事の性質で分類されたものを「カテゴリー」とよぶ

- 名義尺度：「b. 就業状況」，「d. 職種」
  - 数字が分類番号としての意味しか持たない.
    - 「はい」= 1，「いいえ」= 2
  - 他の例. 学籍番号
  - 数字の大きさに本質的な意味がない（数字の大小比較ができない）

- 順序尺度：「c. 仕事の満足度」
  - 数字の順序に意味がある。
  - 他の例. 「1.嫌い，2.普通，3.好き」，震度
  - 数字による<u>大小比較に意味はあるが，計算できない</u>
    - 意味のない計算：「3.好き」 –「1.嫌い」 = 2 ？

量的データ / ニューメリカル・データ

［量的 = Quantitative］数や量で測ったデータ

- 間隔尺度：「g. 婚姻時期」
  - 並び方，値の差には意味がある，値の比率には意味がない。
  - 他の例1. 平成20年生まれは平成10年生まれよりも10歳若いが，2倍若いことを意味しない。
  - 他の例2. 気温 30℃は10℃よりも20℃高いが，3倍暑いことを意味しない。

- 比尺度：「a. 年齢」，「e. 18歳未満の子供の数」，「f. 夫の年収」
  - 並び方，値の差，比率に意味がある。
  - 「年収」900万円は
    - 300万円よりも600万円だけ高く，300万円の3倍である
  - 他の例.　身長，体重，金額，面積，長さ

クロスセクション・データと時系列データ

- クロスセクション・データ（スライド #4）
  - 横断面データともいう。
  - 同じ期間・時期に発生した情報を個体ごと（個人，世帯，企業，地域，物体など）に並べたもの。
  - 並べ方（個体番号）に意味はない。

- 時系列データ
  - 時間の順序にしたがって並べられた情報
  - 時間の単位：年，四半期（3ヶ月），月，週，日，時間，分，秒，・・・

時系列データ（表1.2, p.3）

年次	GDP［兆円］	債務残高［兆円］	消費者物価指数
2001	501.7	582.5	101.5
2002	498.0	643.2	100.6
2003	501.9	670.1	100.3
2004	502.8	751.1	100.3
2005	505.3	813.2	100
2006	509.1	832.3	100.3
2007	513.0	838.0	100.3
2008	489.5	846.7	101.7
2009	473.9	871.5	100.3
2010	479.3	919.2	99.6

出所：内閣府，財務省，総務省統計局

・フローデータ（GDP）

一定期間を単位として，当該時点または期間中に発生した値.

・ストックデータ（債務残高）

過去から蓄積された値.

・指数データ（物価指数）

ある時点の値を基準にして，他の時点の値を基準値に対する比で表わした値.

（例）各年の物価を2005年の価格に対する比で示す．→2005年の価格指数はちょうど100になる.

フローとストック（例.　預金通帳）

日付	お支払い金額	お預かり金額	摘要	残高
07/06	18,773		○○カード	500,000
07/15		200	利息	500,200
08/01		250,000	給与	750,200
08/02	7,540		公共料金	742,660
08/02	210		テスウリヨウ	742,450

フローデータ

ストックデータ
その時点までの量的データの積み重ね

桁区切りと小数点（米英式）

債務残高（10億円）
919,151.1

小数点にドット（or ポイント, ピリオド）
[dot, point, period]

3桁ごとの区切りにカンマ
[Comma]

桁区切りのカンマは桁数が大きいときに利用することが多い.

誤った使い方
0.65%　〇
0,65%　×

質的データの加工（コーディング）(1)

- 文字列を数字に置き変える作業（1 か 0 の値に変換する）
  - 仕事をしている → 1
  - 仕事をしていない→ 0

2値（2項）変数の利点

数字に置き変えることによって，
分析しやすくなる．

表1.3　2値変数

個体番号	就業状況	2値（2項）変数
1	仕事をしている	1
2	仕事をしている	1
3	仕事をしていない	0
4	仕事をしている	1
5	仕事をしている	1
6	仕事をしていない	0
7	仕事をしている	1
8	仕事をしていない	0
9	仕事をしている	1
10	仕事をしている	1
合計		7
平均		0.7

←就業している人の数
←就業している人の割合

質的データの加工（コーディング）(2)

表1.4　職業番号に対応した2値変数

個体番号	d 職業	職種1 専門・技術職	職種2 管理職	職種3 事務職	職種4 販売職	職種5 サービス職
1	1	1	0	0	0	0
2	5	0	0	0	0	1
4	5	0	0	0	0	1
5	2	0	1	0	0	0
7	3	0	0	1	0	0
9	4	0	0	0	1	0
10	5	0	0	0	0	1
合計		1	1	1	1	3
平均		1/7	1/7	1/7	1/7	3/7

$$\text{職種}1 = \begin{cases} 1 & \text{職業が}1\text{のとき} \\ 0 & \text{職業が}1\text{以外のとき} \end{cases} \qquad \text{職種}2 = \begin{cases} 1 & \text{職業が}2\text{のとき} \\ 0 & \text{職業が}2\text{以外のとき} \end{cases}$$

クロス集計表 (1)

- 「仕事をしている」人のうち，「18歳未満の子供がいる」人は何人か？ (表1.7 の C)
- 「仕事をしていない」人のうち，「18歳未満の子供がいない」人は何人か？

表1.7　クロス集計表2

		18歳未満の子供		
		いる	いない	合計
就業状況	仕事をしている	**C = 4**	3	**A = 7**
	仕事をしていない	1	2	3
	合計	**B = 5**	5	**10**

合計人数がわかっていれば，A, B, C を求めるだけで残りは自動的に計算できる。

クロス集計表 (2)

	b	e			
個体番号	就業状況	18歳未満の子供の数	就業状況（仕事をしている=1）	18歳未満の子供の有無（いる場合=1）	仕事をしていて，かつ18歳未満の子供がいる
1	1	0	1	0	0
2	1	2	1	1	1
3	2	0	0	0	0
4	1	2	1	1	1
5	1	1	1	1	1
6	2	0	0	0	0
7	1	0	1	0	0
8	2	3	0	1	0
9	1	0	1	0	0
10	1	1	1	1	1
合計			$A = 7$	$B = 5$	$C = 4$

例題1：表1.13からクロス集計表を作成

ホームページから chap1-econometrics.xlsx をダウンロードして，表1.13のsheet を開く

		雇用形態		
		正規	非正規	合計
最終学歴	a 高校	A		D
	b 専門・短大	B		E
	c 大学・大学院			
	合計	C		25 人

pp.23–27を参照

作成方法1: A,B,C,D,Eに該当する2値変数を作成してクロス集計

1と0からなるの2値変数への変換
= IF (条件式 , 1, 0)

作成方法2: ピボット・テーブルを利用してクロス集計

「挿入」タブ → 「ピボットテーブル」→テーブルまたは範囲を選択

変化率と対数変換 (1) p.8

年次	GDP	対前期**差**	対前期**比**	対前期**変化率**	対前期**変化率** ×100[%]
2001	501.7				
⋮	⋮	⋮	⋮	⋮	⋮
2008	489.5	$489.5-513.0=-23.5$	$\frac{489.5}{513.0}=0.954$	$\frac{489.5-513.0}{513.0}=-0.046$	–4.6%
2009	473.9	$473.9-489.5=-15.6$	$\frac{473.9}{489.5}=0.968$	$\frac{473.9-489.5}{489.5}=-0.032$	–3.2%
2010	479.3	$479.3-473.9=\ \ 5.4$	$\frac{479.3}{473.3}=1.011$	$\frac{479.3-473.3}{473.3}=0.011$	1.1%

例. 2010年のGDPは2009年に比べて,
- ☐ 5.4 兆円増えた
- ☐ 1.011 倍 になった。
- ☐ 1.1 % 変化（成長）した。

chap1-econometrics.xlsx の表1.2のsheet を使って確認

変化率と対数変換 (2) 変化を表す方程式

$$\underset{(変化率)}{1期間の成長率}: \quad g_t = \frac{今期の値 - 前期の値}{前期の値} = \underbrace{\frac{今期の値}{前期の値}}_{対前期比} - 1 \qquad (1.2)式$$

g : growth

今期を第 1 期とすると，前期は第 0 期である。
2010年の GDP を X_1，2009年の GDP を X_0 と書くと，第 1 期の変化率 g_1 は

$$\Rightarrow \quad g_1 = \frac{X_1 - X_0}{X_0} = \underbrace{\frac{X_1}{X_0}}_{対前期比} - 1 \quad \leftrightarrow \quad 1 + g_1 = \frac{X_1}{X_0}$$

$$\leftrightarrow \quad \boxed{X_1 = (1 + g_1) X_0}$$

変化を表す方程式

【重要】今期の GDP (X_1) は前期の GDP (X_0) を **$(1+g_1)$ 倍**した値に等しい

例

a. ギリシャの2016年名目GDPは 175（10億ユーロ）である。ギリシャ政府は来年1年間の目標経済成長率を 2.7% とした。目標となるGDPを計算しなさい。

＝(1+0.027)*175

答え 179.725

b. ある EC サイトの店長は来年の売上を現在よりも 15% 増やすことを目標にしている。今年の売上は600万円であった。目標となる来年の売上を計算しなさい。

＝(1+0.15)*600

答え 690

練習問題 (1)

1. ある会社の今期の売上高は 1153 億円であった。来期の目標売上高を今期よりも 5 %増やしたい。目標売上高を計算しなさい。

2. ある市の現在の人口は 45 万人である。10年後の人口は現在よりも 6% 減ることが予想されている。10年後の人口を計算しなさい。

変化率と対数変換 (3)

変化を表す方程式

$$\boxed{X_1 = (1+g_1)X_0}$$

両辺の対数をとると（底を e [ネイピア数]とする）

$$\leftrightarrow \log_e X_1 = \log_e(1+g_1) + \log_e X_0$$

$$\leftrightarrow \underbrace{\log_e X_1 - \log_e X_0}_{\text{対数差分}} = \log_e(1+g_1)$$

【重要】実は，対数差分 (***LD***; Log Difference) は<u>変化率の近似値</u>であることが知られている。

$$LD_1 = \underbrace{\log_e X_1 - \log_e X_0}_{\text{対数差分}} = \log_e(1+g_1) \approx g_1 \qquad (1.4)\text{式 } p.9$$

近似値（近い値）であることを表す記号（ティルダ）
ただし，g_1が非常に小さい値であることが条件

対数と指数

- 対数とは：指数関数の逆関数をとるための道具
- 指数関数 $y = a^x$ の逆関数（x について解いた式）は指数関数の両辺の対数をとることで得られる。

$$\log_a y = \log_a a^x \rightarrow x = \log_a y$$

- 対数の性質
  I. $\log_a a = 1$
  II. $\log_a xy = \log_a x + \log_a y$
  III. $\log_a (x/y) = \log_a x - \log_a y$
  IV. $\log_a x^y = y \log_a x$
- 底 $a = 10$ の対数を常用対数とよぶ。
  - $\log_{10} 100 = \log_{10} 10^2 = 2$ (∵対数の性質 III)
  - （常用）対数とると桁数がわかる→大きな（or 小さな）数を扱う時に便利

ネイピア数

ネイピア数 e とは

$$\text{十分大きな}\,m\,\text{に対して}\quad e=\left(1+\frac{1}{m}\right)^{m}=2.71828\cdots$$

指数関数の微分の定義からこの形が導かれる

m	$\left(1+\frac{1}{m}\right)^{m}$
10	2.5937425
100	2.7048138
1000	2.7169239
10000	2.7181459
100000	2.7182682
1000000	2.7182805
10000000	2.7182817
100000000	2.7182818
1000000000	2.7182820
10000000000	2.7182821

m が大きくなると，ある一定の値に近づく

ネイピア数 *e* は「自然対数」(*natural logarithm*) の底ともよばれている。
e は対数の底として非常に良く使われるので $\log_e$ のかわりに ln という記号が用いられる。

$$\log_e X=\ln X$$

$$\underbrace{\ln X_1-\ln X_0}_{\text{対数差分}}=\ln(1+g_1)\approx g_1$$

変化率と対数変換 (4)　$\ln(1+g) \approx g$

[%]	変化率		変化率の近似値	誤差
g×100	g	1+g	ln (1+g)	g − ln (1+g)
0.1	0.0010	1.001	0.001000	0.000000
0.2	0.0020	1.002	0.001998	0.000002
0.3	0.0030	1.003	0.002996	0.000004
0.4	0.0040	1.004	0.003992	0.000008
1	0.0100	1.010	0.009950	0.000050
2	0.0200	1.020	0.019803	0.000197
3	0.0300	1.030	0.029559	0.000441
4	0.0400	1.040	0.039221	0.000779
10	0.1000	1.100	0.095310	0.004690
20	0.2000	1.200	0.182322	0.017678
30	0.3000	1.300	0.262364	0.037636
40	0.4000	1.400	0.336472	0.063528

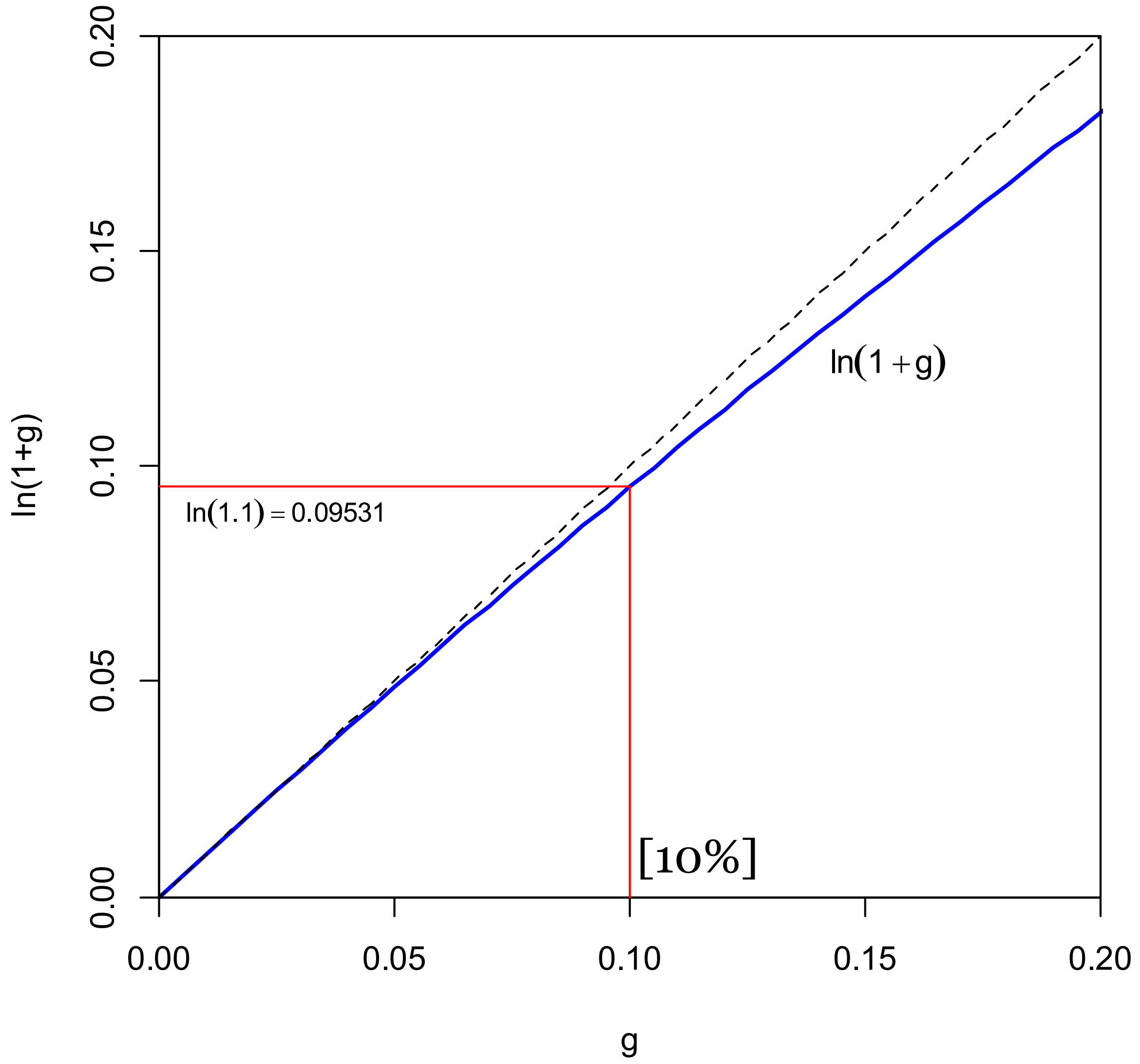
g の値が小さいときは，ln (1+g) は g の近似値になっている。

$$\ln(1+g) \approx g$$

【重要】変化率を測るために，しばしば対数差分 ***LD*** が用いられる。

$$LD_t = \underbrace{\ln X_t - \ln X_{t-1}}_{\text{対数差分}} \approx g_t$$

g と $\ln(1+g)$ の違い



0.20
0.15
0.10
0.05
0.00
ln(1+g)
ln(1 + g)
ln(1.1) = 0.09531
[10%]
0.00
0.05
0.10
0.15
0.20
g

変化率と対数変換 (5)

- Excel関数
  - 対数 **=log(数値，底)**
  - 自然対数 **=ln(数値)**
  - ネイピア数のべき乗 **=exp(数値)**
    - =exp(1) と入力すると $e = 2.71828\ldots$が計算できる

変化率と対数変換 (6)

表1.6, p.8

		対数	GDP変化率	対数差分 LD_t
年次	X_t (GDP)	$\ln X_t$	$g_t = \dfrac{X_t - X_{t-1}}{X_{t-1}}$	$\ln X_t - \ln X_{t-1} \approx g_t$
2001	501.7	6.2180	.	.
2002	498.0	6.2106	–0.0074	–0.0074
2003	501.9	6.2184	0.0078	0.0078
2004	502.8	6.2202	0.0018	0.0018
2005	505.3	6.2252	0.0050	0.0050
2006	509.1	6.2326	0.0075	0.0075
2007	513.0	6.2403	0.0077	0.0076
2008	489.5	6.1934	–0.0458	–0.0469
2009	473.9	6.1610	–0.0319	–0.0324
2010	479.3	6.1723	0.0114	0.0113

例題2（pp.27—30）も参照

変化率と対数変換 (7)

対数差分から正確な変化率を計算し直す方法

$$\boxed{LD_t} = \underbrace{\ln X_t - \ln X_{t-1}}_{\text{対数差分}} = \boxed{\ln(1+g_t)} \approx g_t$$

対数の計算ルール

$$\boxed{\begin{array}{l} \log_a a^2 = 2 \\ \log_e e^x = x \end{array}}$$

$$LD_t = \ln(1+g_t)$$
$$\leftrightarrow e^{LD} = 1 + g_t$$
$$\leftrightarrow g_t = e^{LD} - 1$$

ネイピア数を *LD* 乗した値から1
を引くと，ちょうど*g* になる

表1.6 の場合

=exp(GDP対数差分)－1 と入力すると，GDP変化率と同じ値が得られる。

例. 2010 年の GDP 対数差分 LD = 0.0013

→ exp(0.0013) −1 = 0.011394809

まとめ

- データには質的データと量的データがある。
- 質的データには名義尺度と順序尺度がある。
- 量的データには間隔尺度と比尺度がある。
- 質的データを0か1に変換して2値変数を作成すると，クロス集計の役に立つ。
- 時系列データの「変化」を測る方法として，差，比，変化率がある。
- 変化率 = 比 − 1 となる。
- ある1期間の変化率を g とおくとき，今期の値は前期の値を $1 + g$ 倍したものに等しい。
- 対数差分は変化率の近似値になる（変化率が小さい場合）。したがって，変化を表す指標として対数差分はたいへん良く用いられる。