経済統計a:第七回

担当教員　黒田敏史

今週の内容

- テキスト6章「データの身近な分析方法」
  - 代表値とその他の記述統計
  - 変化率と構成比
  - 弾力性と寄与度

今週の内容

- テキスト6章「データの身近な分析方法」
  - 代表値とその他の記述統計
  - 変化率と構成比
  - 弾力性と寄与度

代表値とその他の記述統計

- 代表値
  - 集団から得られたデータを一つの値で表現する値
    - 平均･中央値･分散等、記述統計とも呼ばれる
    - どの数値で代表値を作成するかは分析目的に依存
      - GDPは国の経済的な豊かさを表現するが、人口の多寡が反映されないため、国民の豊かさを反映しない（購買力平価で表現した中国のGDPは7,096B$は日本の4,297B$より大きい）
    - どの代表値を用いる事が適切であるかは、分析目的に依存
      - 例：集団の傾向に偏りが無い場合に全体の傾向を知るなら平均が望ましいが、データが極端に偏っているときは中央値の方が望ましいときがある
    - 代表値をどの範囲の標本から取得するかも、分析目的に依存
      - 例：日本国民の所得の不平等を計測するとき、日本国民全体の分散を用いると、同じ年齢の人の間の不平等や、貧しい人の間の不平等についての情報を得ることが出来ない

代表値とその他の記述統計

- 平均に関する代表値
  - 相加平均：N個の数を足し合わせて、標本数で割る
    - 相加平均(mean) $= X_1 + X_2 + \cdots + X_N = \left(\sum_{i=1}^{N} X_i\right) / N$
  - 加重平均: N個の数に重みwを掛けたものを足し合わせる。重みは$\sum_{i=1}^{N} w_i = 1$ とする。
    - 加重平均(weighted aveage) $= w_1 X_1 + w_2 X_2 + \cdots + w_N X_N = \sum_{i=1}^{N} w_i X_i$
    - 相加平均・加重平均は共に算術平均に含まれる
  - 幾何平均: N個の数を掛け合わせてN乗根をとる
    - 幾何平均(geometric mean) $= X_1 \times X_2 \times \cdots \times X_N = \prod_{i=1}^{N} X_1$

代表値とその他の記述統計

- 平均に関する代表値
  - 調和平均:Xの逆数の算術平均の逆数
    - 調和平均(harmonic mean) $= N \bigg/ \sum_{i=1}^{N_1} \frac{1}{X_i}$
    - 例:片道100kmの道のりを行きは時速10km、帰りは時速100kmで走った場合をトータルした平均時速は、200km/[(100/10)+(100/100)時間]=18.19km
分母分子を100で割ると上述の幾何平均の式になり、2/(1/10+1/100)＝18.19が得られる
    - 例2:株式購入のドル=コスト平均法では、毎月特定銘柄を一定金額購入する。ある銘柄を毎月20万円購入する場合、先月の株価が5万円、今月の株価が4万円の時の平均購入額は調和平均となり、4.4万円。一定数量を購入していた場合の平均購入額は4.5万円となり、後者の方が割高になる。

代表値とその他の記述統計

- 3つの平均値の関係
  - 常に相加平均≧相乗平均≧調和平均
  - 等しくなるのは、全ての標本の値が等しいとき

代表値とその他の記述統計

- 平均に関する代表値
  - 平方平均：Xの２乗を算術平均した結果の平方根
    - 平方平均(squared mean) = $\sqrt{\left(\sum_{i=1}^{N_1} X_i^2\right) \Big/ N}$
  - 中位値（中央値）
    - Xを大きさの順に並べたときに、中央に来る値のこと
    - 奇数（odd)の場合、N=2k+1とすると、中央値$(median) = X_{k+1}$
    - 偶数（odd)の場合、N=2kとすると、中央値 $= \left(X_k + X_{k+1}\right) / 2$

代表値とその他の記述統計

- 平均に関する代表値
  - 最頻値：N個のデータの中で最も数の多い値
    - 例：６面体サイコロを降って以下のような目が出た場合の最頻値は5

1回目	2回目	3回目	4回目	5回目	6回目	7回目	8回目	9回目	10回目
3	5	2	1	5	3	2	5	6	1

  - トリム平均：N個のデータのうち、上下の何%を標本から取り除いて計算した平均値
    - スポーツの採点等に用いられている

代表値とその他の記述統計

- 移動平均
  - 移動平均とは、毎期観察されるデータの性質を得るために用いられており、平均を計算する標本の範囲を一定の期間に区切ったもののことを言う
  - 3期移動平均: $3期移動平均 = \left(X_{t-1} + X_t + X_{t+1}\right)/3$
  - 5期移動平均: $5期移動平均 = \left(X_{t-2} + X_{t-1} + X_t + X_{t+1} + X_{t+2}\right)/5$
  - 偶数の場合は移動平均の中心化を行う

$$4期移動平均 = \frac{\left(X_{t-2} + X_{t-1} + X_t + X_{t+1}\right)/4 + \left(X_{t-1} + X_t + X_{t+1} + X_{t+2}\right)/4}{2}$$

代表値とその他の記述統計

- 移動平均
  - 後方移動平均：基準時点より前の一定期間を用いて計算を行う

    後方3期移動平均 $=\left(X_{t-2}+X_{t-1}+X_{t}\right)/3$

  - 反復移動平均：基準時点の前後の移動平均をさらに平均する

    3期反復移動平均

$$=\frac{\left(X_{t-2}+X_{t-1}+X_{t}\right)/3+\left(X_{t-1}+X_{t}+X_{t+1}\right)/3+\left(X_{t}+X_{t+1}+X_{t+2}\right)/3}{3}$$

$$=\frac{X_{t-2}+2X_{t-1}+3X_{t}+2X_{t+1}+X_{t+2}}{3}$$

代表値とその他の記述統計

- 散らばりに関する代表値
  - 分布範囲：最大値と最小値の差 $R = X_{\max} - X_{\min}$
    - 最大と最小の間の散らばりに一切反応しない
  - 四分位偏差：第3四分位数と第1四分位数の差の1／2
    $R = \left(Q_3 - Q_1\right)/2$
    平均偏差： $MD = \left(\sum \left|X_i - M\right|\right)/N$
  - 標準偏差： $SD = \sqrt{\sum \left(X_i - M\right)^2 / N}$
    - 分散：標準偏差の2乗 $V = SD^2 = \sum \left(X_i - M\right)^2 / N$
    - 与件が一定ならば、相対的に貧しい人から相対的に豊かな人への移転において必ず分散を大きくする
    - 平均が大きいと、分散も大きくなる

代表値とその他の記述統計

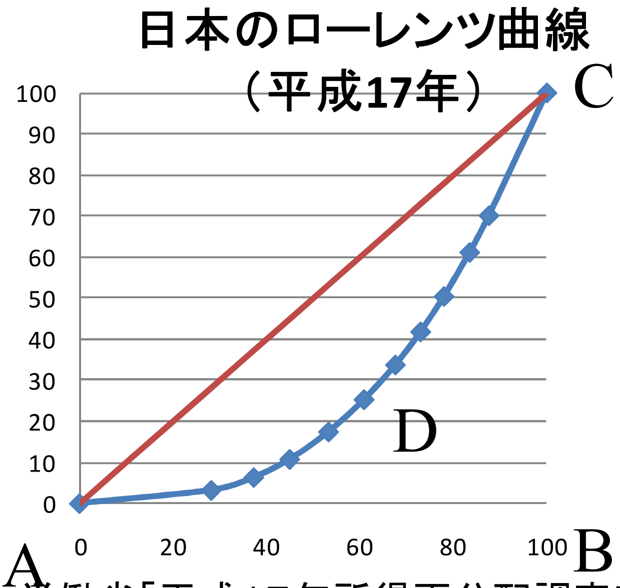
- 不平等度の指標
  - 相対平均偏差：平均偏差を平均値で割る
    完全平等で0，完全不平等で（２－２／N)

    $$D_e = \left(\sum \left| X_i - M \right|\right) \Big/ NM$$

    - 平均より上の人、もしくは下の人同士の所得移転に反応しない
  - 変動係数：標準偏差を平均値で割る
    完全平等で0、完全不平等で $\sqrt{N-1}$
    - 指数が平均と独立するが、同じ所得の移転の大きさであれば、指数がどこから何処への移転であるかに依存しない

代表値とその他の記述統計

- 不平等度の指標
  - ローレンツ曲線
    - 所得の一番低い人から高い人に順に並べ、累積所得を縦軸にプロットしてゆく
      - 完全平等なら直線AC
      - 完全不平等ならABC
    - 45度線と曲線の間の面積の大きさが不平等の尺度となるため、視覚的に不平等を表現



出典：厚生労働省「平成１７年所得再分配調査報告書」

代表値とその他の記述統計

- 不平等度の指標
  - ジニ係数
    - 定義：N人から作成される任意のペアの所得の差の絶対値を足し合わせ、$2MN^2$ で割る $G = \frac{1}{2MN^2}\sum_{j=1}^{N}\sum_{i=1}^{N}\left|y_i - y_j\right|$
    - 完全不平等で1、完全平等で0をとる
    - ローレンツ曲線を用いて計算することも可能

$$\frac{ACD - ABC}{ABC} = 1 - \frac{\sum_{i=1}^{N} r_i (y_i - y_{i-1})}{10,000}$$

r_iはi番目に低い所得水準の構成比

  - ジニ係数の所得要素による分解
    - 所得要素k毎に計算された擬ジニ係数を所得要素のシェアで加重平均したものは、ジニ係数と一致する

代表値とその他の記述統計

- 日本のジニ係数の推移

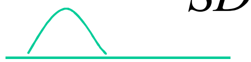
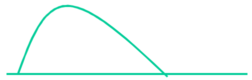
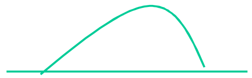
	等価当初所得	① ＋社会保障給付金 ー社会保険料	等価可処分所得 （②ー税金）	等価再配分所得 ③＋現物給付	再配分による改善度	社会保障による改善度
	①	②	③	④		
1993年	0.3703	0.3313	0.3097	0.3074	17.00%	11.20%
1996年	0.3764	0.3273	0.3119	0.3096	17.70%	13.70%
1999年	0.4075	0.3501	0.3372	0.3326	18.40%	15.30%
2002年	0.4194	0.3371	0.3227	0.3217	23.30%	19.90%
2005年	0.4354	0.3355	0.3218	0.3225	25.90%	22.80%

※１再分配による改善度＝１ー④／①
※２社会保障による改善度＝１ー②／①×④／③
※３税による改善度＝１ー③／②
注：平成11年以前の現物給付は医療のみ
平成14年以降については医療、介護、保育
出典：厚生労働省「平成１７年所得再分配調査報告書」

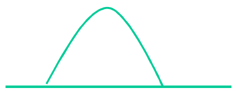
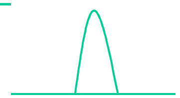
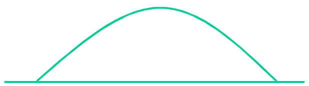
代表値とその他の記述統計

- 不平等度の指標
  - ローレンツ曲線・ジニ係数の特徴
    - 全てのペアの差から指数を作成すると言う意味で、不平等に関する多くの情報を用いている
    - しかし、ある移転を評価する時の敏感さは、その移転が行われる個人間にどれだけの個人がいるかに依存してしまう
      - 平均所得が異なる2つのグループで比較を行う際に、平均所得の多寡が考慮されない
      - いかなる個人がどの所得水準にあるのかを考慮しない
      - 人数が異なる2つのグループを比較する際、人数の多寡を考慮しない
      - 異なる2つのローレンツ曲線が交差した場合に、どちらが好ましいかを比較できない
      - 不平等度に関するより詳細な研究については、アマルティア・セン『不平等の経済学』鈴村興太朗・須賀晃一訳、東洋経済新報社、2000.等を参照。

代表値とその他の記述統計

- その他の代表値
  - 歪度：分布の偏り（非対称度）を表す $S_k = \dfrac{\left(\sum_1^N (X_i - M)^3\right)/N}{SD^3}$
    - $S_k = 0$：左右対称、平均＝中央値＝最頻値 
    - $S_k > 0$：右側の方向き、最頻値＜中央値＜平均 
    - $S_k < 0$：左側に傾き、平均＜中位値＜最頻値 
  - 尖度：分布が正規分布に比べて平坦か、尖っているかを表す

$$K = \frac{\left(\sum_1^N (X_i - M)^4\right)/N}{SD^4}$$

    - K=3のとき、正規分布と同程度 
    - K<3のとき、正規分布よりも尖っている 
    - K>3のとき、正規分布よりも平坦 

代表値とその他の記述統計

- 偏差値
  - ある標本の値が集団の中でいかなる位置にあるかを表した指数

$$SS = \frac{10(X_i - M)}{SD} + 50$$

  - X_iが平均と一致していればSS=0,平均以下なら50以下、平均以上なら50以上をとる

今週の内容

- テキスト6章「データの身近な分析方法」
  - 代表値
  - 変化率と構成比
  - 弾力性と寄与度

変化率と構成比

- 基準指数と接続指数
  - 基準指数:データの分析において、時系列的な変化を問題にする場合には、特定時点を100とした相対的な値を用いる方が便利な場合がある
    - t年におけるt-n年を基準とした基準指数: $y_t^{t-n} = \frac{Y_t}{Y_{t-n}}$
    - 5年後に基準指数を再計算する場合、$y_{t-n-x}^{t-n} = \frac{y_{t-n}^{t-n}}{y_{t-n}^{t-n-5}} \times y_{t-n-x}^{t-n-5}$

変化率と構成比

- 変化率・瞬間風速
  - 変化率:ある基準時点から別の時点へとどれだけ変化したかを表す
    - t-n年からt年にかけての変化率: $R_Y = \dfrac{Y_t - Y_{t-n}}{Y_{t-nt}} \times 100$
  - 年平均変化率(瞬間風速):計測可能な長さが異なる場合、それを1期間に平準化する
    - t-n年からt年にかけての変化率: $R_Y = \left(\left(\dfrac{Y_t}{Y_{t-n}} \right)^{\frac{1}{n}} - 1 \right) \times 100$
    - 1年より短い期間を1年に延長する場合を、瞬間風速という

変化率と構成比

- 構成比・発生率・特化係数
  - 構成比：特定集団内の個別要素iの占める割合

$$C_i = \frac{J_i}{\sum_i J} \times 100$$

  - 発生率：特定集団において要素がiである確率（いわゆるシェア）

$$P_i = \frac{N_i}{\sum_i N_i} \times 100$$

  - 特化係数：特定集団内jで発生する要素iが、他の集団で発生する要素iに比べて頻度が高いか否かを記述する

$$S_i = \frac{J_{ij}}{\sum_j J_{ij}} \frac{\sum_i J_{ij}}{\sum_j \sum_i J_{ij}}$$

今週の内容

- テキスト6章「データの身近な分析方法」
  - 代表値
  - 変化率と構成比
  - 弾力性と寄与度

弾力性と寄与度

- 弾力性
  - ある変数Aが変化したとき、それに伴って他の変数Bが何%変化するかを表す
    - 例:需用の価格弾力性、需要の所得弾力性、供給の価格弾力性
    - 点弾力性:価格の対前年度比で需要量の対前年度比を割った値 $$\varepsilon = \frac{D_t - D_{t-1}}{D_t} \div \frac{P_t - P_{t-1}}{P_t}$$
    - 平均値周りの弾力性:価格・需要量の平均値と平均値周りでの需要曲線の傾きbから計算される弾力性 $$\varepsilon = \frac{\Delta D}{D} \div \frac{\Delta P}{P} = \frac{\partial D}{\partial P}\frac{\overline{P}}{\overline{D}}$$

弾力性と寄与度

- 対数と弾力性
  - 変数yが変数xの関数で表現される場合 $y = f(x)$
  - 両辺に対数をとる $\log y = \log f(x)$
  - 左辺をxで微分すると、 $\dfrac{d \log y}{dy} = \dfrac{1}{y}$
  - 右辺をxで微分すると $\dfrac{d \log x}{dx} = \dfrac{1}{x}$
  - よって、 $\dfrac{d \log y}{d \log x} = \dfrac{dy}{dx}\dfrac{x}{y} = \varepsilon$ となる

弾力性と寄与度

- 寄与度・寄与率
  - あるデータが複数のデータを合成して作成されているとき、あるデータの変化を要素となる複数のデータの変化率に分解する事が出来る
    - 例：Y=C+I+Gの場合

$$dY = dC\frac{\partial Y}{\partial C} + dI\frac{\partial Y}{\partial I} + dG\frac{\partial Y}{\partial G}$$

$\frac{\partial Y}{\partial C} = \frac{\partial Y}{\partial I} = \frac{\partial Y}{\partial G} = 1$ であるから、$dY = dC + dI + dG$

両辺Yで割って、 $\frac{dY}{Y} = \frac{C}{Y}\frac{dC}{C} + \frac{I}{Y}\frac{dI}{I} + \frac{G}{Y}\frac{dG}{G}$

    - この各項を寄与度と呼ぶ

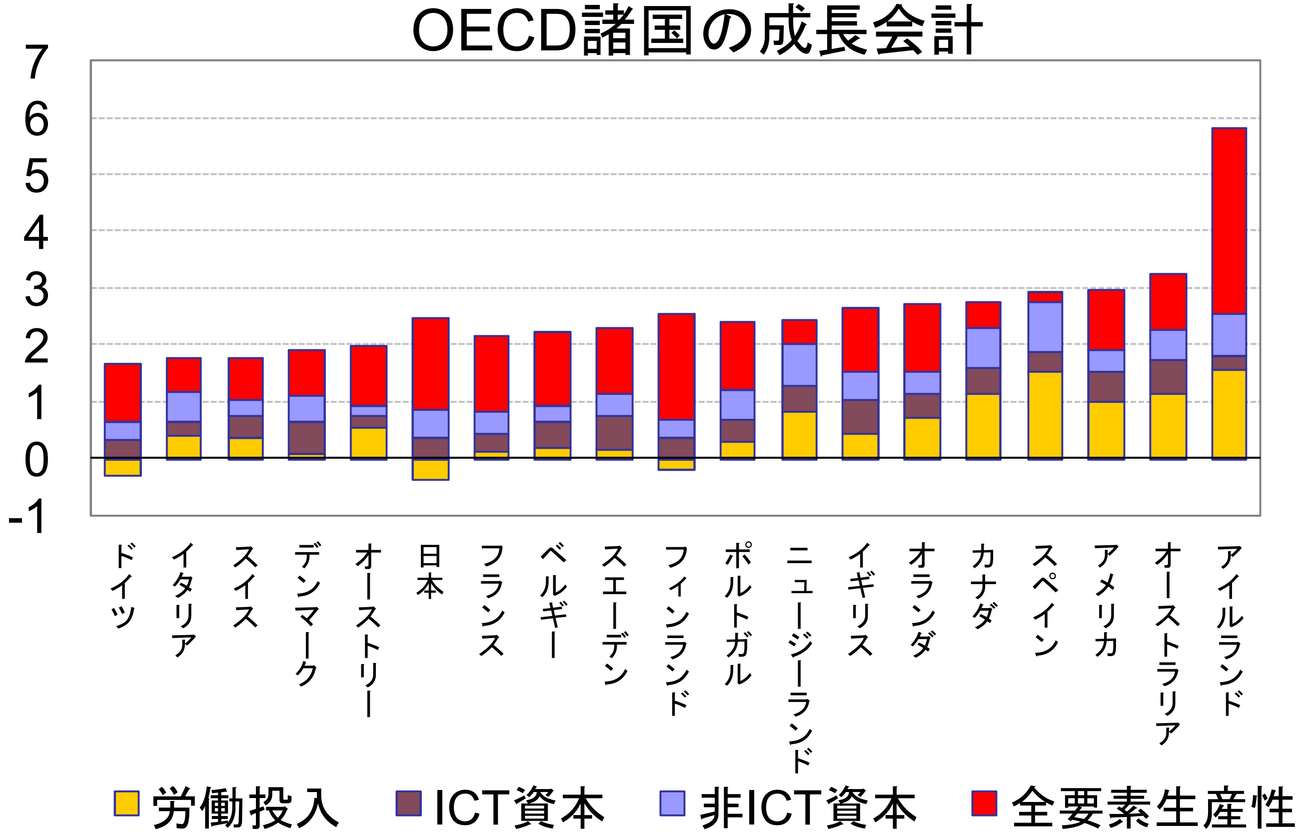
弾力性と寄与度

- 寄与度・寄与率
  - ある寄与度 $\frac{X}{Y}\frac{dX}{X}$ を CD_x と置く
  - このとき、変化率の各項目別の寄与度を寄与率と呼ぶ

    寄与率 $= CR_x = \frac{CD_x}{R_y} \times 100$

弾力性と寄与度

- OECD諸国の成長会計
  - 以下の図はOECD諸国の1985年-2006年の経済成長を、労働投入の増加・ICT資本の増加・非ICT資本の増加・全要素生産性に分解した図



出典:OECD Factbook 2009

今週の内容

- テキスト6章「データの身近な分析方法」
  - 代表値とその他の記述統計
  - 変化率と構成比
  - 弾力性と寄与度

来週の内容

- グラフと印象操作
  - グラフはデータを直感的に把握可能にする
  - しかし、正しくグラフを作成しなければ、データの正しい理解の妨げになる事もある
  - しばしば意図的に誤った理解を促す事を意図したと思われるグラフが用いられる事がある
  - 幾つかのケースを用いて、正しい作成方法と、誤ったグラフの発見方法について取り扱う