

# 第3回：記述統計の確認

北村 友宏

2020年5月22日

# 本日の内容

1. gretl での記述統計の出力
2. 記述統計表の作成

# 記述統計

- ▶ データセットを読み込んだ gretl の画面上で、記述統計を出力したい変数を選択し、右クリック→「基本統計量」と操作し、Show **full** statistics を選んで「OK」をクリックすると、選んだ変数の様々な統計量が表示される。

# Show full statistics での統計量

- ▶ 平均

- ▶  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$

- ▶ 中央値

- ▶ 観測値を小さい順に並べたときに中央に来る値.
  - ▶ 観測値数  $n$  が偶数の場合は中央で隣り合う2つの値の平均値.

- ▶ 標準偏差

- ▶  $s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.$

- ▶ 最小値

- ▶  $\min\{x_i\}.$

- ▶ 最大値

- ▶  $\max\{x_i\}.$

▶ 変動係数

▶  $CV_x = \frac{S_x}{\bar{x}}$ .

▶ 歪度

- ▶ (学部レベルを超えるので省略)

▶ 過剰尖度

- ▶ (学部レベルを超えるので省略)

▶  $\alpha$  百分位数

- ▶ 観測値を小さい順に並べたときに  $0.01\alpha n$  番目に来る値.
- ▶ 50 百分位数は中央値と同じ.
- ▶ gretl では 5 百分位数と 95 百分位数が出力できる.

- ▶ IQ range: 四分位範囲 (Interquartile range)
  - ▶ 75 百分位数 – 25 百分位数.
- ▶ 欠損値数
  - ▶ 値が観測されていない個体（データセットの Excel ファイルにおいて、その変数のセルが空白になっている個体）の数.
  - ▶ e.g., この授業の実習で用いている都道府県別・男女別の所得と消費のデータでは、富山県の男性の所得と消費が欠損.

# 実習 1

1. gretl を起動.
2. 「ファイル」 → 「データを開く」 → 「ユーザー・ファイル」と操作.
3. 消費 2009.gdt を選択し, 「開く」をクリック.
4. 「income」から「consumption\_th」までの6つをドラッグして選択し, その上で右クリック → 「基本統計量」と操作.
5. Show **full** statistics を選んで「OK」をクリックすると, 選択した変数の記述統計 12 種類が表示される.

	平均	中央値	最小値	最大値
income	2.3356e+005	2.2771e+005	1.4464e+005	4.5307e+005
consumption	1.8202e+005	1.7795e+005	1.1270e+005	3.1978e+005
male	0.50000	0.50000	0.00000	1.0000
female	0.50000	0.50000	0.00000	1.0000
income_th	233.56	227.71	144.64	453.07
consumption_th	182.02	177.94	112.70	319.78

  

	標準偏差	変動係数	歪度	過剰尖度
income	52722	0.22574	1.2904	2.5538
consumption	37459	0.20580	0.94532	1.5378
male	0.50268	1.0054	0.00000	-2.0000
female	0.50268	1.0054	0.00000	-2.0000
income_th	52.722	0.22574	1.2904	2.5538
consumption_th	37.459	0.20580	0.94532	1.5378

  

	5百分位数	95百分位数	IQ range	欠損値数
income	1.6923e+005	3.4066e+005	63032	1
consumption	1.2890e+005	2.5833e+005	45091	1
male	0.00000	1.0000	1.0000	0
female	0.00000	1.0000	1.0000	0
income_th	169.23	340.66	63.031	1
consumption_th	128.90	258.33	45.091	1

このような画面が表示されれば成功。

Mac の PC では、小数点以下の表示桁数が異なっている場合がある。

まだ作業があるので、「gretl: 基本統計量」のウィンドウは**まだ閉じない!**



- ▶ 統計量の名前の位置がズレていることに注意.
- ▶ e+005 は,  $\times 10^5$  という意味.
  - ▶ e.g., 変数 income (円単位の可処分所得) の平均は  $2.336 \times 10^5$  (円).

まだ作業があるので, 「gretl: 基本統計量」のウィンドウは**まだ閉じない!**

6. 表示されている記述統計の画面上で右クリック→「名前を付けて保存...」と操作.
7. 出てきたダイアログボックスの、「標準テキスト」を選び、「OK」をクリック.
8. 記述統計 5 月 22 日.txt という名前で「2020 ミクロデータ分析 1」フォルダに保存. 次の実習で使うので、「gretl: 基本統計量」のウィンドウは**まだ閉じない!**

# レポートや論文での記述統計表の作成

見やすく，理解しやすい表を載せるには，

- ▶ 表番号と表のタイトルをつける。
- ▶ 変数名は統計解析ソフトでの変数名そのままではなく，分かりやすいように書き直す。
- ▶ 単位がある変数は単位を明記する。
- ▶ 小数の数値はあまり細かく表示せず，小数第1～4位程度まで示せば十分。縦方向に見たとき，小数点の位置が揃うようにする。
- ▶ 観測値数が全変数について同じ場合，表の下に「注（Note）」として「観測値数は\*\*\*である」などと注記してもよい。

この授業では Word を用いた記述統計表の作成方法を説明する。

## 実習 2

1. Word を起動し，記述統計 5 月 22 日.docx という名前で「2020 ミクロデータ分析 1」フォルダに保存.
2. 「挿入」→「表」と操作して 3 行 6 列の表を作る.
3. 表全体をドラッグし，「参照設定」→「図表番号の挿入」と操作.
4. ラベルを「表」に，位置を「選択した項目の上」して OK をクリックすると，表のすぐ上の行に「表 1」と入力される.
  - ▶ ラベルに「表」がなければ，「新しいラベル...」をクリックして出てくるダイアログボックスの入力ボックスに表と入力して OK をクリック.
5. 「表 1」の後に全角スペースを入れて記述統計と入力し，中央揃えにする.

6. 表の 1 行 2 列目に平均値，1 行 3 列目に標準偏差，1 行 4 列目に最小値，1 行 5 列目に最大値，1 行 6 列目に中央値と入力.
7. 表の 1 行 2 列目から 1 行 6 列目までをドラッグし，「レイアウト」タブ（右端の，色が濃いほう）から「配置」→「中央揃え」と操作.
8. 表の 2 行 1 列目に消費支出（千円），3 行 1 列目に可処分所得（千円）と入力.
9. 表の 2 行 1 列目から 3 行 1 列目までをドラッグし，「レイアウト」タブ（右端の，色が濃いほう）から「配置」→「中央揃え」と操作.

10. gretl で出力された表の中で， consumption\_th と income\_th の各統計量の数値を， Word で作成した表の対応するセルにコピー・貼り付けする． 数値をドラッグして選択し， 右クリック→「コピー」と操作すればコピーできる．
- ▶ consumption\_th は消費支出（千円）， income\_th は可処分所得（千円）．
  - ▶ 数値は必要に応じて四捨五入して各列で小数点以下の表示桁数を統一し， 縦方向に見たとき， 小数点の位置が揃うようにする． 今回は， 標準偏差の数値について， 小数第 3 位を四捨五入する．
11. 表の 2 行 2 列目から 6 行 3 列目まで（数値の入っているセル）をドラッグし， 「レイアウト」タブ（右端の， 色が濃いほう）から「配置」→「中央揃え（右）」と操作．

12. Word で作成した表のすぐ下の行に，（注）観測値数は 93 である．と入力して上書き保存．

- ▶ 観測値数は，gretl の画面下部の「日付なし：全範囲 1 - 94」と書かれている部分を見れば分かる．ただし，記述統計に出力された通り，consumption\_th や income\_th に欠損値が 1 個あるので，この場合の観測値数は  $94 - 1 = 93$  となる．

表 1 記述統計

	平均値	標準偏差	最小値	最大値	中央値
消費支出 (千円)	182.02	37.46	112.70	319.78	177.94
可処分所得 (千円)	233.56	52.72	144.64	453.07	227.71

(注) 観測値数は 93 である。

このような表を作成できればよい。



# 作成した記述統計表から分かること

- ▶ 可処分所得の平均値よりも消費支出の平均値が小さい。  
⇒ 所得の一部が消費に回される傾向がある。
- ▶ 両変数とも標準偏差は平均値に対して極端に大きくない。  
⇒ 都道府県や男女によるブレ幅が過度に大きくない。
- ▶ 消費支出に関しては、極端に大きな、または極端に小さな値（異常値, outlier）はない。
- ▶ 両変数とも平均値と中央値が近い。  
⇒ 平均値は分布の偏りや異常値に、過度に引っ張られていない。

本日の作業はここまで.

今回は gretl のデータセットに変更を加えていない  
ので, **gretl のデータセット (消費 2009.gdt)** を上書き  
保存する必要はない.