

# 第 11 回：データセット作成の 復習

北村 友宏

2020 年 12 月 18 日

# 本日の内容

1. データの入手・Excelでの加工と整理
2. gretlでの取り込み
3. 記述統計の出力

# 今後の授業の内容

- ▶ 回帰モデルの被説明変数がダミー変数の場合の分析方法の解説と実習
- ▶ データ：日本プロサッカーリーグ（Jリーグ）『出場記録』の，2011年におけるヴィッセル神戸チーム所属選手の年間試合出場時間と総得点，および翌年にチームを移籍したかどうか（選手別データ）
- ▶ 分析：「サッカー選手のチーム移籍に影響を与える要因」の分析
  - ➡ 推定するモデルの被説明変数がダミー変数（移籍する = 1，移籍しない = 0）
- ▶ 参考：鹿野繁樹（2015）『新しい計量経済学—データで因果関係に迫る』日本評論社.

# 実習 1

「サッカー選手のチーム移籍に影響を与える要因」を分析するためのデータを入手し、まずは Excel の加工・整理を行う。

1. Excel を起動し、jleaguekobe2011.xlsx という名前で 2020 ミクロデータ分析 2 フォルダに保存。
2. Jリーグ公式 HP (<https://www.jleague.jp/>) にアクセス。
3. 「成績・データ」→「データ集」→「出場記録」→「2013年以前はこちら」→「2011年：」の「Jリーグ ディビジョン1」の「ヴィッセル神戸」と辿る。

4. 表の 7 行 1 列目 (No) 以下全体 (最終行の右から 3 列目まで) をコピーし、  
jleaguekobe2011.xlsx のセル A1 を選択して貼り付け。Excel ファイルの 1 行目が全て空白になった場合は、1 行目を削除。
5. C 列を選択し、新たな列を挿入。
  - ▶ B 列の右隣に列を挿入。
6. セル C1 に、Transfer と入力。
7. セル B1 に入力されている「選手」を消去し、Player と入力。
8. セル D1 に入力されている「出場」を消去し、Attend と入力。
9. セル E1 に入力されている「時間」を消去し、Time と入力。
10. セル F1 に入力されている「得点」を消去し、Goal と入力。

11. G列からX列までを選択し、右クリック→「削除」と操作して、G列からX列までの全てのセルに入力された情報を消去。
  - ▶ 今回の分析では、G列以降の部分は不要。
12. 「選手出場記録」が開かれているブラウザの戻るボタンをクリック。
13. 「2012年：」の「Jリーグ ディビジョン1」の「ヴィッセル神戸」をクリック。

14. 表示された 2012 年の選手名と、  
jleaguekobe2011.xlsx に貼り付けた 2011 年の  
選手名 (B 列) を照合して、  
2011 年に在籍した選手が **2012 年にも在籍し  
ていれば** (2012 年のデータにも氏名があれ  
ば), jleaguekobe2011.xlsx の C 列の該当選手  
のセルに **0** を入力。  
2011 年に在籍した選手が **2012 年には在籍して  
いなければ** (2012 年のデータには氏名がなけ  
れば), jleaguekobe2011.xlsx の C 列の該当選  
手のセルに **1** を入力。
- ▶ 翌年 (2012 年) に移籍すれば 1, しなければ 0 と  
するダミー変数を作成。
  - ▶ ここでは、移籍の定義を「翌年には当該チームに  
在籍していないこと」とする (他チームへの異動  
や選手を辞めることなども含めて)。

## 15. jleaguekobe2011.xlsx を上書き保存して閉じる.

- ▶ 作成した Excel ファイルの B 列 (変数名 : Player) の選手名が全角日本語になっているが, 現行バージョンの gretl では変数の観測値に全角日本語が使われていても文字化けせずに読み込める.



## 実習 2

続いて、実習 1 で作成した Excel ファイルを gretl で取り込む。

1. gretl を起動。
2. jleaguekobe2011.xlsx を、gretl の画面にドラッグ・アンド・ドロップ。
3. 出てきたダイアログボックスの、インポートを開始する場所: の列: と行: がともに 1 になっていることを確認し、「OK」をクリックすると、データが読み込まれる。
4. 「インポートされたデータは・・・(中略)・・・解釈し直しますか?」というメッセージが表示されるので、「いいえ」をクリック。
  - ▶ 作成した jleaguekobe2011 は複数個体 (選手) ・単一時点 (2011 年のみ) の**横断面データ**。

5. 「No」から「Goal」までの6つをドラッグして選択し、その上で右クリック→「データ（値）を表示」と操作すると、全変数の観測値リストが新規ウィンドウにて表示される。

gretl: データ表示

No	Player	Transfer	Attend	Time
1	紀氏 隆秀	1	0	0
2	近藤 岳登	0	19	1625
3	相馬 崇人	0	8	545
4	北本 久仁衛	0	34	3060
5	河本 裕之	0	32	2880
6	羽田 憲司	0	6	217
7	朴 康造	0	27	2030
8	松岡 亮輔	1	22	1831
9	ホジェリニョ	1	11	637
10	ボッティ	1	26	1774
11	ボボ	1	32	2104
12	大久保 嘉人	0	30	2395
13	宮本 恒靖	1	4	164
14	大屋 翼	0	2	180
15	吉田 孝行	0	29	1700
16	田中 英雄	0	26	2115
17	有田 光希	1	10	170
18	森岡 亮太	0	21	588
19	茂木 弘	0	27	2285
20	馬場 賢治	1	1	21
21	イジミン	1	3	110
22	三原 俊	0	18	1404
23	石 雅	1	16	1402
24	三島 祐平	1	4	17
25	都倉 賢	0	14	568
26	井住	0	2	121

このような画面が表示されれば成功. onekr の観測値リストは, 下のほうに表示されている. 確認したら閉じる.

先ほど出力した観測地リストは、Transfer以降の変数の観測値がズレて表示され見づらいが、これは観測値に全角日本語を含む変数（Player）を表示させているため。

6. Ctrl キーを押しながら「No」「Transfer」「Attend」「Time」「Goal」の5つを左クリックして選択し、その上で右クリック→「データ（値）を表示」と操作すると、Playerを除く5変数の観測値リストが新規ウィンドウにて表示される。

The screenshot shows a window titled "gretl: データ表示" (gretl: Data Display). The window contains a table with 28 rows and 5 columns. The columns are labeled "No", "Transfer", "Attend", "Time", and "Goal". The data is as follows:

	No	Transfer	Attend	Time	Goal
1	1	1	0	0	0
2	2	0	19	1625	0
3	3	0	8	545	1
4	4	0	34	3060	2
5	5	0	32	2880	0
6	6	0	6	217	0
7	7	0	27	2030	5
8	8	1	22	1831	2
9	9	1	11	637	0
10	10	1	26	1774	1
11	11	1	32	2104	7
12	13	0	30	2395	9
13	14	1	4	164	0
14	15	0	2	180	0
15	17	0	29	1700	9
16	18	0	26	2115	2
17	19	1	10	170	0
18	20	0	21	588	2
19	21	0	27	2285	1
20	22	1	1	21	0
21	23	1	3	110	0
22	24	0	18	1404	0
23	25	1	16	1402	1
24	26	1	4	17	0
25	27	0	14	568	2
26	28	0	2	121	0

このような画面が表示されれば成功。観測値リストが見やすく表示された。確認したら閉じる。

※もし数字が違っていたら、データセット (jleaguekobe2011.xlsx) の作成の際にミスをしているということなので、データセットの作成からやり直すこと。

7. メニューバーから「ファイル」→「データに名前を付けて保存」と操作し、jleaguekobe2011.gdt という名前で「2020 ミクロデータ分析 2」フォルダに保存。

先ほどの実習でダウンロードし，加工・整理した日本プロサッカーリーグ（Jリーグ）『出場記録』の，2011年におけるヴィッセル神戸チーム所属選手の年間試合出場時間と総得点，および翌年にチームを移籍したかどうか（選手別データ）のデータは，

- ▶ 横断面データ
- ▶ ミクロデータ

# 各変数の定義

データセットの計 6 変数の定義は、以下の通り。

- ▶ No : 選手番号
- ▶ Player : 選手名
- ▶ Transfer : 移籍ダミー
  - ▶ 翌年 (2012 年) に移籍した = 1
  - ▶ 翌年 (2012 年) に移籍しなかった (残留した) = 0
- ▶ Attend : 年間出場試合数
- ▶ Time : 年間出場時間
- ▶ Goal : 年間総得点



## 実習 3

1. 「Transfer」から「Goal」までの3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.
2. 「主要な統計量を表示する」が選ばれている状態で「OK」をクリックすると，選択した変数の記述統計5種類が表示される.
  - ▶ 古くても 2020年8月6日版では，この表示が日本語化されている.

	平均	中央値	標準偏差	最小値	最大値
Transfer	0.4242	0.0000	0.5019	0.0000	1.000
Attend	14.33	12.00	12.14	0.0000	34.00
Time	1020	588.0	1039	0.0000	3060
Goal	1.333	0.0000	2.508	0.0000	9.000

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.

古くても 2020 年 8 月 6 日版では，上の画像のように統計量名が全て日本語で表示される.

- ▶ 統計量の名前の位置がズレていて見づらいが、各変数について出力された数字は左から平均，中央値，標準偏差，最小値，最大値の順.

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！**

3. 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.
4. 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.
5. 記述統計 12月18日.txt という名前で「2020 ミクロデータ分析 2」フォルダに保存. 本日の作業はここまで.