

## 中田研究室 3 年生ゼミ データ分析演習

## 1. 目的

実際のサッカーデータについて、R プログラムにより、様々な情報を表示できるようにする。それを通して、データ分析技術の基礎を身に着ける。

## 2. グループ課題

## 2.1. 基本的なデータ洗浄などの処理を済ませたデータを保存したファイル

”data\_sample\_clean3.csv”を以下のコマンドを用いて R で読み込んで確認せよ。

```
data <- read.csv("data_sample_clean3.csv", head=TRUE)
```

なお、読み込むデータは、1 レコード (1 行) が図 1 のような構成となっており、文字列の 2 次元配列 data に格納されている。追跡人物情報部は、29 個の追跡人物情報で構成されており、各々の追跡人物情報は、表 1 に示す 7 つの値で構成されている。ボール情報は、表 2 に示す 7 つの値で構成されている。また、イベント情報は、表 3 に示す 2 つの値で構成されている。表 4 には、イベント情報の中で用いられるイベント番号を示している。なお、これらの情報で欠損値には”NA”が入っている。

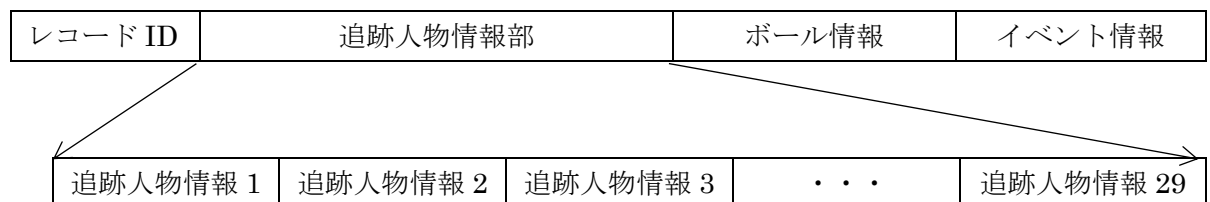


図 1. 読み込んでいるデータの 1 レコードの構成

表 1. 追跡人物情報の構成

値	説明
チーム ID	赤 (左) チームプレイヤー : 0, 赤 (左) チームキーパー : 10, 青 (右) チームプレイヤー : 1, 青 (右) チームキーパー : 11, 審判 : 3 (その他の数字が用いられることもあるが、基本的に無視して良い.)
追跡人物 ID	データスタジアム社の選手追跡システムが利用する ID (基本的に気にしなくて良い)
ゼッケン	選手のゼッケンを表す. 審判や試合に参加していない選手については - 1 が入る.

x 座標	グラウンド上での x 座標（中心を 0 とした横方向．右が正）を表す．単位は cm である．
y 座標	グラウンド上での y 座標（中心を 0 とした縦方向．奥が正）を表す．単位は cm である．
x 方向速度	グラウンド上での x 方向の速度（右が正）を表す．単位は cm/秒であり，1 秒前のレコード x 座標との差分により算出している．
y 方向速度	グラウンド上での y 方向の速度（奥が正）を表す．単位は cm/秒であり，1 秒前のレコード y 座標との差分により算出している．

表 2. ボール情報の構成

値	説明
x 座標	グラウンド上での x 座標（中心を 0 とした横方向．右が正）を表す．単位は cm である．
y 座標	グラウンド上での y 座標（中心を 0 とした縦方向．右が正）を表す．単位は cm である．
z 座標	グラウンド上での y 座標（中心を 0 とした上下方向．上が正）を表す．単位は cm である．
x 方向速度	グラウンド上での x 方向の速度（右が正）を表す．単位は cm/秒であり，1 秒前のレコード x 座標との差分により算出している．
y 方向速度	グラウンド上での y 方向の速度（奥が正）を表す．単位は cm/秒であり，1 秒前のレコード y 座標との差分により算出している．
z 方向速度	グラウンド上での z 方向の速度（上が正）を表す．単位は cm/秒であり，1 秒前のレコード z 座標との差分により算出している．
ボール保持チーム	ボールを保持しているチーム． 赤（左）チーム：0，青（右）チーム：1，
ボール状態	ボールの状態．Alive：0，Dead：1

表 3. イベント情報の構成

値	説明
チーム ID	赤（左）チームプレイヤー：0，青（右）チームプレイヤー：1 ※なお，イベント無しの場合，0が入っている．
イベント番号	発生イベント（の一部）の種類を表す 0～8 の数字が入っている．

表 4. イベント情報に格納されているイベント番号

イベント番号	イベントの説明
0	イベント無し
1	キックオフ
2	コーナーキック
3	ゴールキック
4	ペナルティキック
5	間接フリーキック
6	直接フリーキック
7	シュート
8	オフサイド
9	ファウル
10	スローイン

## 2.2. 基本的なデータ洗浄などの処理を済ませたデータを保存したファイル

”data\_sample\_clean3.csv”をファイル”data\_processing.R”に記載の R プログラムコードを用いて、基本的なフォーメーション特徴量を格納した加工後データに変換せよ。

なお、図 1 は、加工後データの 1 レコードの構成を表している。両チーム（非キーパー）情報、左チーム（非キーパー）情報、右チーム（非キーパー）情報は、各々、表 5 に示すような 11 個の値で構成されている。左チームキーパー情報、右チームキーパー情報、審判情報は、各々、表 6 に示すような 4 つの値で構成されている。ボール情報は、表 7 に示すような 6 つの値で構成されている。イベント情報は加工前と同様に表 4 に示すような構成となる。

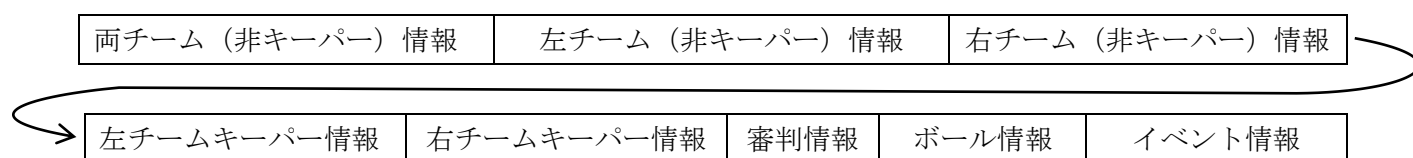


図 1. 加工後データの 1 レコードの構成

表 5. 加工後データにおけるチーム（非キーパー）情報の構成

値	説明
x 平均値	グラウンド上での x 座標（中心を 0 とした横方向．右が正）の平均値である．単位は cm である．
x 最小値	グラウンド上での x 座標（中心を 0 とした横方向．右が正）の最小値である．単位は cm である．
x 最大値	グラウンド上での x 座標（中心を 0 とした横方向．右が正）の最大値である．単位は cm である．
x 標準偏差	グラウンド上での x 座標（中心を 0 とした横方向．右が正）の標準偏差である．単位は cm である．
x 範囲	グラウンド上での x 座標（中心を 0 とした横方向．右が正）の範囲（最大値－最小値）である．単位は cm である．
y 平均値	グラウンド上での y 座標（中心を 0 とした縦方向．奥が正）の平均値である．単位は cm である．
y 最小値	グラウンド上での y 座標（中心を 0 とした縦方向．奥が正）の最小値である．単位は cm である．
y 最大値	グラウンド上での y 座標（中心を 0 とした縦方向．奥が正）の最大値である．単位は cm である．
y 標準偏差	グラウンド上での y 座標（中心を 0 とした縦方向．奥が正）の標準偏差である．単位は cm である．
y 範囲	グラウンド上での y 座標（中心を 0 とした縦方向．奥が正）の範囲（最大値－最小値）である．単位は cm である．
xy 相関係数	グラウンド上での xy 座標の（ピアソンの積率）相関係数である．単位は存在せず-1～1 の値を取る．
vx 平均値	グラウンド上での x 方向（中心を 0 とした横方向．右が正）の速度の平均値である．単位は cm/秒である．
vx 最小値	グラウンド上での x 方向（中心を 0 とした横方向．右が正）の速度の最小値である．単位は cm/秒である．
vx 最大値	グラウンド上での x 方向（中心を 0 とした横方向．右が正）の速度の最大値である．単位は cm/秒である．
vx 標準偏差	グラウンド上での x 方向（中心を 0 とした横方向．右が正）の速度の標準偏差である．単位は cm/秒である．
vx 範囲	グラウンド上での x 方向（中心を 0 とした横方向．右が正）の速度の範囲（最大値－最小値）である．単位は cm/秒である．
vy 平均値	グラウンド上での y 方向（中心を 0 とした縦方向．奥が正）の速度の平均値である．単位は cm/秒である．

vy 最小値	グラウンド上での y 方向（中心を 0 とした縦方向．奥が正）の速度の最小値である．単位は cm である．
vy 最大値	グラウンド上での y 方向（中心を 0 とした縦方向．奥が正）の速度の最大値である．単位は cm である．
vy 標準偏差	グラウンド上での y 方向（中心を 0 とした縦方向．奥が正）の速度の標準偏差である．単位は cm である．
vy 範囲	グラウンド上での y 方向（中心を 0 とした縦方向．奥が正）の速度の（最大値－最小値）である．単位は cm である．
vxvy 相関係数	グラウンド上での x y 方向の速度間の（ピアソンの積率）相関係数である．単位は存在せず-1～1 の値を取る．

表 6. 加工後データにおけるキーパー・審判情報の構成

値	説明
x 座標	グラウンド上での x 座標（中心を 0 とした横方向．右が正）である．単位は cm である．
y 座標	グラウンド上での y 座標（中心を 0 とした縦方向．奥が正）である．単位は cm である．
x 方向速度	グラウンド上での x 方向（中心を 0 とした横方向．右が正）の速度である．単位は cm/秒である．
y 方向速度	グラウンド上での y 方向（中心を 0 とした縦方向．奥が正）の速度である．単位は cm である．

表 7. 加工後データにおけるボール情報の構成

値	説明
x 座標	グラウンド上での x 座標（中心を 0 とした横方向．右が正）を表す．単位は cm である．
y 座標	グラウンド上での y 座標（中心を 0 とした縦方向．右が正）を表す．単位は cm である．
z 座標	グラウンド上での y 座標（中心を 0 とした上下方向．上が正）を表す．単位は cm である．
x 方向速度	グラウンド上での x 方向の速度（右が正）を表す．単位は cm/秒であり，1 秒前のレコード x 座標との差分により算出している．
y 方向速度	グラウンド上での y 方向の速度（奥が正）を表す．単位は cm/秒であり，1 秒前のレコード y 座標との差分により算出している．
z 方向速度	グラウンド上での z 方向の速度（上が正）を表す．単位は cm/秒であり，1 秒前のレコード z 座標との差分により算出している．

**2.3.** 加工後のデータを用いて、全ての位置情報に関連する値について、時系列グラフを作成せよ。なお、ファイル”time\_series\_plot.R”に記載の例を参考とせよ。また、映像データやイベント情報と照らし合わせ、考察せよ。ただし、横軸はフレーム数であり、1 秒間 25 フレームである。

**2.4.** 加工後のデータを用いて、全ての位置情報に関連する値について、イベント毎の箱ひげ図を作成せよ。なお、ファイル”box\_plot.R”に記載の例を参考とせよ。また、イベント毎にどのような特徴があるか考察せよ。

### 3. 個別課題

- ・ グループで実施した分析を基に以下のような分析を実施せよ。
  - ・ 速度情報に関連する値についての時系列グラフ・箱ひげ図の作成と考察
  - ・ 特定のいくつかの値についてのイベント毎のヒストグラムの作成と考察
  - ・ 特定のいくつかの値について試合の対称性を考慮した箱ひげ図の作成と考察 (180 度の回転対称性,  $x$  軸に対する対称性,  $y$  軸に対する対称性)。
- ・ (オプション) 更により高度な分析を実施せよ。
  - ・ 主成分分析, クラスタ分析など教師無しの分析をし, その結果を考察。
  - ・ 決定木学習, ロジスティック回帰分析などで特定のイベントを予測するモデルを作成し, その結果を考察。

### 4. 発表について

3 週目に発表する。