

2016/11/29 Tue.

問題解決技法入門

3. Data Analysis

2. Data Visualization using R

堀田 敬介

Rでデータの視覚化

- csv ファイルをデータとして利用
 - 「マイドキュメント(Y:)」に「R」フォルダをつくり中に保存

bb2016.csv

※) 2016年プロ野球セ・パ成績 (Yahoo Japan! Sports naviより)

	リーグ	試合数	勝数	負数	引分数	勝率	得点	失点	本塁打	盗塁	打率	防御率
広島	セ	143	89	52	2	0.631	684	497	153	118	0.272	3.2
巨人	セ	143	71	69	3	0.507	519	543	128	62	0.251	3.45
DeNA	セ	143	69	71	3	0.493	572	588	140	67	0.249	3.76
阪神	セ	143	64	76	3	0.457	506	546	90	59	0.245	3.38
ヤクルト	セ	143	64	78	1	0.451	594	694	113	82	0.256	4.73
中日	セ	143	58	82	3	0.414	500	573	89	60	0.245	3.65
日本ハム	パ	143	87	53	3	0.621	619	467	121	132	0.266	3.06
ソフトバンク	パ	143	83	54	6	0.606	637	479	114	107	0.261	3.09
ロッテ	パ	143	72	68	3	0.514	583	582	80	77	0.256	3.66
西武	パ	143	64	76	3	0.457	619	618	128	97	0.264	3.85
楽天	パ	143	62	78	3	0.443	544	654	101	56	0.257	4.11
オリックス	パ	143	57	83	3	0.407	499	635	84	104	0.253	4.18

- ファイルの読み込み

※1行目にheaderあり

※各行の名称は列1に

```
> dfbb <- read.csv("Y:/R/bb2016.csv", header=T, row.names=1)
```

※ファイルのフルパス

例) YドライブのRフォルダ内にあるbb2015.csvという名前のファイル

Rでデータの視覚化

- 読みデータの確認
 - dfbbに代入したdata frame の中身を**全て**表示

```
> dfbb
```
 - dfbbに代入したdata frame の中身を**一部(先頭)**表示

```
> head(dfbb)
```
 - dfbbに代入したdata frame の中身を**一部(後尾)**表示

```
> tail(dfbb)
```
 - dfbbの**項目名**表示 (header=Tで読んだデータ)

```
> names(dfbb)
```
 - dfbbの**レコード名**表示 (row.names=1で指定した)

```
> row.names(dfbb)
```

Rでデータの視覚化

- 箱ひげ図を描画

※dfbb\$本塁打 ... data.frameである dfbbの項目“本塁打”を箱ひげ図のデータとして使用

```
> boxplot(dfbb$本塁打)
```

... ①
- オプションを指定し箱ひげ図を描画

```
> boxplot(dfbb$本塁打, col="tomato", xlab="本塁打", ylab="本数", main="12チーム本塁打数の箱ひげ図")
```

... ②

<オプション>
col ... 色の指定(colour)
xlab ... x軸のラベル(label)
ylab ... y軸のラベル(label)
main ... タイトル

①

②

12チーム本塁打数の箱ひげ図

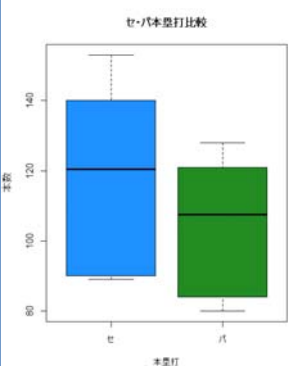
Rでデータの視覚化

- グループ毎に箱ひげ図を描画

```
> boxplot(dfbb$本塁打~dfbb$リーグ, xlab="本塁打", ylab="本数",  
col=c("dodgerblue", "forestgreen"), main="セ・パ本塁打比較")
```

※「リーグ(セ・パ)」毎に描画するよう指定
(tilde(~)で層別描画を指定している)

<オプション>
col ... 色の指定(colour)
c("blue", "red", "green") ... 色名のベクトルをつくる
xlab ... x軸のラベル(label)
ylab ... y軸のラベル(label)
main ... タイトル



Rでデータの視覚化

- 幹葉図(stem-and-leaf plot)を描画

```
> stem(dfbb$本塁打)
```

The decimal point is 1 digit(s) to the right of the |
8 | 0490
10 | 134
12 | 188
14 | 03

- 幹葉図を描画(オプション scale=2)

```
> stem(dfbb$本塁打, 2)
```

The decimal point is 1 digit(s) to the right of the |
8 | 049
9 | 0
10 | 1
11 | 34
12 | 188
13 |
14 | 0
15 | 3

※scale数を大きくするとより詳細な幹葉図に
(default=1)

Rでデータの視覚化

- csv ファイルをデータとして利用

– 「マイドキュメント(Y:)」に「R」フォルダをつくり中に保存
bi2016.csv

氏名	チーム	リーグ	打率	試合数	打数	安打	二塁打	三塁打	本塁打	打点	得点	盗塁	犠打	犠打率	出塁率	長打率	打点率	盗塁率	盗塁数					
坂本 勇人	阪	セ	0.344	137	576	488	168	28	3	23	271	75	96	67	81	0	1	6	13	0.433	0.555	0.339	4	16
鈴木 誠也	阪	セ	0.335	129	528	466	156	26	6	29	285	95	76	79	53	3	3	16	0.404	0.612	0.346	10	2	
岡倉 勇貴	D	セ	0.322	133	561	469	151	28	4	44	319	110	88	105	87	3	0	2	0	0.43	0.68	0.393	6	2
豊田 泰介	阪	セ	0.315	141	640	574	181	23	3	13	248	96	92	106	46	0	2	3	13	0.398	0.432	0.343	3	4
梅原 孝介	神	セ	0.313	131	523	455	142	23	3	11	205	59	51	78	43	3	4	4	0	0.393	0.453	0.31	4	1
山田 哲人	ヤ	セ	0.308	135	596	481	144	26	3	38	292	102	101	97	8	6	4	30	0.425	0.607	0.299	14	5	
村田 修一	阪	セ	0.3024	143	576	529	160	31	0	25	267	81	98	83	38	9	2	2	1	0.354	0.505	0.305	23	15
川崎 慎吾	ヤ	セ	0.3023	103	458	420	127	23	1	1	154	32	48	31	34	1	1	2	3	0.354	0.367	0.301	13	5
前井 貴浩	阪	セ	0.3	132	513	454	135	23	2	19	220	101	66	101	54	1	0	4	0	0.372	0.485	0.323	12	5

※) 2016年プロ野球個人成績 (Yahoo Japan! Sports naviより)

- ファイル読み込み

```
> dfbi <- read.csv("Y:/R/i2016.csv", header=T, row.names=1)
```

【演習】
好きな項目を1つ選び(例: 打率, 安打, 本塁打, 打点, 得点, etc.), 12チーム毎の箱ひげ図をつくれ. オプションとして, 色, x軸ラベル, y軸ラベル, タイトルを適切に指定すること

参考文献

- 山本他「Rで学ぶデータサイエンス12統計データの視覚化」共立出版(2013)
- 青木繁伸『Rによる統計解析』オーム社(2009)
- 荒木孝治『RとRコマンダーではじめる多変量解析』日科技連(2007)
- 金明哲『Rによるデータサイエンス』森北出版(2007)
- 新納浩幸『Rで学ぶクラスター解析』オーム社(2007)

もっと知りたい人へ

- 関連する経営学科の授業
 - 「統計の見方」(1/2セメ)
 - 「統計の分析と利用」(2セメ)
 - 「データ処理Ⅱ」(2セメ)
 - 「統計データの扱い方」(3/4セメ)
 - 「多変量の統計データ解析」(4セメ)

Rでデータの視覚化

その他のグラフ作成例

棒グラフ

散布図

※これらのグラフを作成したい時は、Excelを使った方が良い

Rでデータの視覚化

棒グラフを作成

※色指定用のベクトル生成。"royalblue"を6回 repeat し、
"violetred"を6回repeat したベクトルをつくり cc に代入

> cc <- c(rep("royalblue",6), rep("violetred",6))

> barplot(dfbb\$勝数, names.arg=row.names(dfbb), col=cc, xlab="チーム名", ylab="勝数")

dfbb\$勝数 ... data.frameである dfbb の項目 "勝数"を棒グラフのデータとして使用

names.arg ... それぞれの棒に対応する名称

col ... 棒の色指定

xlab ... x軸のラベル

ylab ... y軸のラベル

Tips !

> colors()

※Rで使える657色の
の名称リスト表示

Rでデータの視覚化

散布図を作成(1)

> plot(dfbb\$勝率, dfbb\$防御率, xlab="勝率", ylab="防御率", col="purple")

x軸を dfbb\$勝率

y軸を dfbb\$防御率

のデータを用い散布図を作成

xlab ... x軸ラベルの指定

ylab ... y軸ラベルの指定

col ... プロットする点の色指定

dfbb\$勝率 は dfbb[,6] でもよい

dfbb\$防御率 は dfbb[,12] でもよい

Rでデータの視覚化

散布図を作成(2)

> plot(dfbb[,6], dfbb[,12], xlab="勝率", ylab="防御率", type="b")

x軸を dfbb[,6]="勝率"

y軸を dfbb[,12]="防御率"

のデータを用い散布図を作成

xlab ... x軸ラベルの指定

ylab ... y軸ラベルの指定

type ... 描画点の種類

"p" ... points 点 (default)

"l" ... lines 線分

"b" ... both点と線分 両方

"c" ... "b" から点を抜いたもの

"o" ... overplotted

"h" ... histogram

"s" ... stair steps

"n" ... no plotting 点をかかない

Rでデータの視覚化

- 散布図を作成(3) ※プロットはせずに、枠・軸だけを描画

```
> plot(dfbb[,6], dfbb[,12], xlab="勝率", ylab="防御率", type="n")
> text(dfbb[,6], dfbb[,12], dfbb[,1], col=cc)
```

※リーグ名称をプロット点として描く
(data.frameであるdfbbの1列目に
リーグ名を入れたことを思いだそう！)

※col=ccは色設定をccにすること
(ccはリーグ毎の色設定用ベクトル
として作ったことを思いだそう！)

dfbb[,1]はdfbb\$リーグでもよい
dfbb[,6]はdfbb\$勝率でもよい
dfbb[,12]はdfbb\$防御率でもよい

Rでデータの視覚化

- 散布図を作成(4) ※プロットはせずに、枠・軸だけを描画

```
> plot(dfbb[,6], dfbb[,12], xlab="勝率", ylab="防御率", type="n")
> text(dfbb[,6], dfbb[,12], row.names(dfbb))
```

※チーム名称をプロット点としてかく
(read.csvでcsvファイルを読み込んだ時に、
row.namesとして1列目のチーム名称を指定したことを思いだそう！)

dfbb[,6]はdfbb\$勝率でもよい
dfbb[,12]はdfbb\$防御率でもよい

【参考】 Rでデータの視覚化

- 箱ひげ図と散布図を作成(1) -scatterplot()-

```
> install.packages("car")
> library(car)
```

※scatterplot()の使用準備
package "car"のインストール
package "car"の読み込み

```
> scatterplot(dfbb[,4], dfbb[,8], xlab="負数", ylab="失点")
```

x軸をdfbb[,4]="負数"
y軸をdfbb[,8]="失点"
のデータを用い散布図を作成

xlab ... x軸ラベルの指定
ylab ... y軸ラベルの指定

※それぞれの軸に、それぞれのデータの箱ひげ図が描かれる

※緑線は回帰直線 regression line

※赤線・赤点線は平滑化線とspan

【参考】 Rでデータの視覚化

- 箱ひげ図と散布図を作成(2) -scatterplot()-

```
> install.packages("sp")
> install.packages("maptools")
> library(sp)
> library(maptools)
```

※pointLabel()の使用準備
packages "sp","maptools"のインストール
packages "sp","maptools"の読み込み
(注:必ずsp→maptoolsの順！)

– 点とチーム名を両方プロットする

```
> scatterplot(dfbb[,4], dfbb[,8], xlab="負数", ylab="失点", reg.line=F, smooth=F)
> pointLabel(x=dfbb[,4], y=dfbb[,8], labels=row.names(dfbb))
```

※平滑化線は描かない

※散布図の点のラベルをrow.names(dfbb)として書く

※回帰直線 regression lineは描かない(FはFalseの意)

