

Fusion 補遺ーデータ解析環境 R 入門

関西大学商学部 荒木孝治

2019/10/30

- ・データ解析環境 R : 統計学・機械学習
- ・Python : ディープラーニング・機械学習

1. R 入門

- (1) 電卓的使い方ー演算子 (足し算+, 引き算-, かけ算*, 割り算/)

```
> 1 + 2  
[1] 3
```

- (2) 変数の利用ーx、y

```
> x <- 10  
> y <- 20
```

- (3) 変数の足し算

```
> z = x + y
```

- (4) 結果の表示

```
> z  
[1] 30  
> (z = x * y)
```

- (5) 計算と同時に結果を表示ー () で囲む

```
[1] 200
```

- (6) ベクトルの定義

```
> v1 <- c(1, 2, 3, 2, 1)  
> v2 <- c(10, 20, 30, 40, 50)
```

(7) ベクトルの足し算—対応する位置の数字同士の足し算

```
> (plus_v1v2 = v1 + v2)
[1] 11 22 33 42 51
```

(8) ベクトルの結合 (column: 列)。他に rbind()がある (row: 行)

```
> (set_v1v2 = cbind(v1, v2))
      v1 v2
[1,]  1 10
[2,]  2 20
[3,]  3 30
[4,]  2 40
[5,]  1 50
```

(9) 平均の計算—関数 mean()

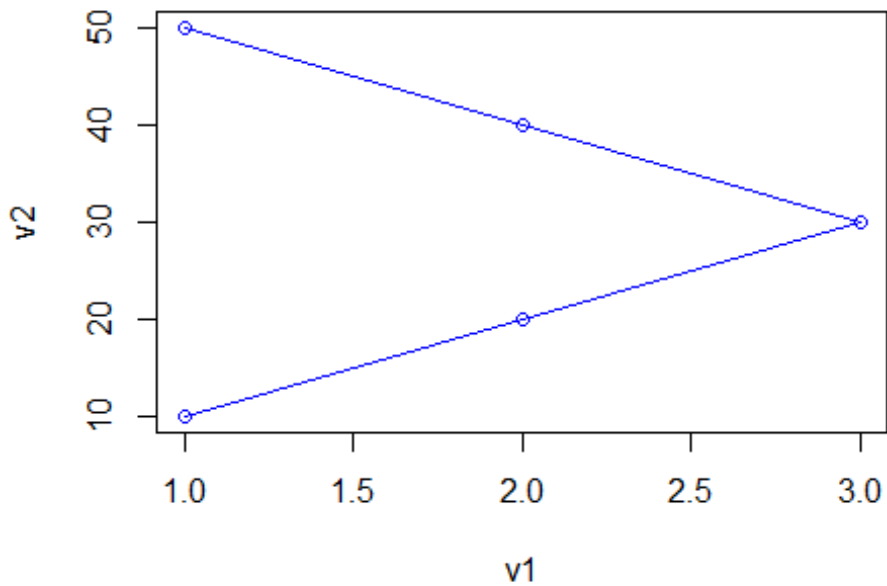
```
> mean(plus_v1v2)
[1] 31.8
```

(10) データの要約情報 (最小値、第 1 四分位数、メディアン、第 3 四分位数、最大値)

```
> summary(plus_v1v2)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  11.0   22.0   33.0   31.8   42.0   51.0
```

(11) 行列のプロット (第 1 列を横軸、第 2 列を縦軸)

```
> plot(set_v1v2, type = "o", col = "blue")
```



(12) エクセルファイルの読み込み。関数 `library()` を用いてパッケージ `readxl` を起動

```
> library(readxl)
```

(13) 関数 `read_excel()` の利用。パス・ファイル名を指定

```
> tokyo_drink <- read_excel("C:/data/tokyo_drink.xlsx")
```

(14) 読み込まれたデータセット `tokyo_drink` をコピー (`td`)

```
> td <- tokyo_drink
```

(15) 関数 `lm()` を用いて回帰分析の実行

モデルの書式：目的変数～説明変数 1+説明変数 2+……

```
> lm_result <- lm(td$飲料販売量 ~ td$気温 + td$湿度 + td$土日祝ダミー)
```

結果の表示：関数 `summary()` の利用

```
> summary(lm_result)
```

Call:

```
lm(formula = td$飲料販売量 ~ td$気温 + td$湿度 + td$土日祝ダミー)
```

Residuals:

Min	1Q	Median	3Q	Max
-142.718	-49.742	5.141	46.590	127.925

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	374.9633	14.5686	25.738	<2e-16 ***
td\$気温	5.8803	0.4914	11.965	<2e-16 ***
td\$湿度	0.4903	0.2318	2.115	0.0351 *
td\$土日祝ダミー	-86.7386	6.9676	-12.449	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 62.69 on 362 degrees of freedom
 Multiple R-squared: 0.5183, Adjusted R-squared: 0.5143
 F-statistic: 129.8 on 3 and 362 DF, p-value: < 2.2e-16

(16) 別の回帰モデルの作成 `lm_res2`, `lm_res3`

```
> lm_res2 <- lm(td$飲料販売量 ~ td$気温 + td$土日祝ダミー)
> lm_res3 <- lm(td$飲料販売量 ~ td$湿度 + td$土日祝ダミー)
```

(17) パッケージ `memisc` の起動

```
> library(memisc)
Loading required package: lattice
Loading required package: MASS

Attaching package: 'memisc'

The following object is masked from 'package:car':
  recode

The following objects are masked from 'package:stats':
  contr.sum, contr.treatment, contrasts

The following object is masked from 'package:base':
  as.array
```

(18) 関数 `mtable()` を利用して、複数の回帰分析の結果の比較

```
> mtable(lm_result, lm_res2, lm_res3)
```

Calls:

```
lm_result: lm(formula = td$飲料販売量 ~ td$気温 + td$湿度 + td$土日祝ダミー)
```

lm_res2: lm(formula = td\$飲料販売量 ~ td\$気温 + td\$土日祝ダミー)
 lm_res3: lm(formula = td\$飲料販売量 ~ td\$湿度 + td\$土日祝ダミー)

```

=====
                                lm_result    lm_res2    lm_res3
-----
(Intercept)    374.963***    400.563***    376.813***
                (14.569)      (8.151)      (17.185)
td$気温        5.880***      6.394***
                (0.491)      (0.429)
td$湿度        0.490*
                (0.232)
td$土日祝ダミー -86.739***    -87.746***    -84.283***
                (6.968)      (6.984)      (8.216)
-----
R-squared      0.518          0.512          0.328
N              366          366          366
=====
Significance: *** = p < 0.001; ** = p < 0.01;
              * = p < 0.05
  
```

2. 電力消費データの分析

```
> library(readxl)
```

(19) エクセルデータの読み込み

```
> elec_weather <- read_excel("C:/data/elec_weather.xlsx")
```

(20) 切り出す変数の指定

```
> lm_name <- c("年月日", "月", "平日 or 土日祝", "時", "東京電力 (1 時間単  
位)", "東京_気温 (1 時間単位)",  
+ "東京_降水量 (1 日合計値)")
```

(21) 必要なデータの切り出し

```
> elwe <- elec_weather[, lm_name]
```

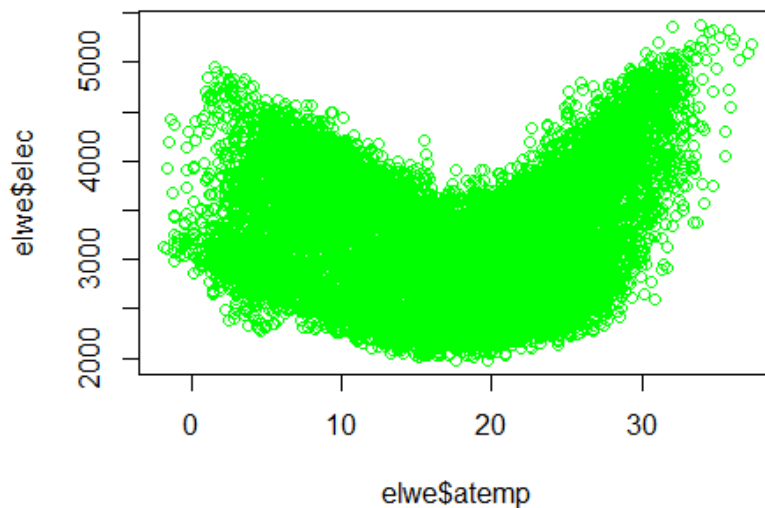
(22) 変数名の変更

```
> names(elwe) <- c("ymd", "month", "hei_dns", "hour", "elec", "atemp",  
"wdrop")
```

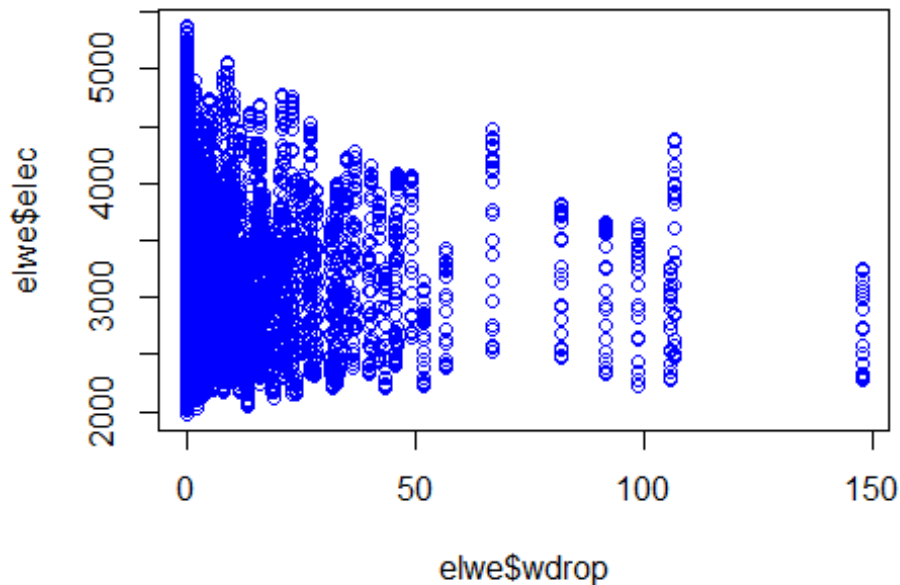
(23) 散布図の作成 (plot(横軸に取る変数, 縦軸に取る変数))

Col (color: 色の指定)

```
> plot(elwe$atemp, elwe$elec, col = "green")
```



```
> plot(elwe$wdrop, elwe$elec, col = "blue")
```



(24) 変数 atemp の 2 乗の変数 atemp2 の追加

```
> elwe$atemp2 <- elwe$atemp^2
```

(25) 変数 atemp の 3 乗の変数 atemp3 の追加

```
> elwe$atemp3 <- elwe$atemp^3
```

(26) ダミー変数 wadr_dummy の作成 (雨量 wdrop が正(>0)のとき、1、他のとき、0)

```
> elwe$wadr_dummy <- ifelse(elwe$wdrop > 0, 1, 0)
```

(27) 回帰分析 (2 乗、3 乗の項、ダミー変数の利用)

```
> lm_re1 <- lm(elwe$elec ~ elwe$atemp + elwe$atemp2 + elwe$atemp3 + elwe$wadr_dummy)
```

(28) モデル lm_re1 の回帰係数の表示 (関数 coef()の利用)

```
> coef(lm_re1)
```

(Intercept)	elwe\$atemp	elwe\$atemp2	elwe\$atemp3
3453.6410544	35.4280133	-8.5273962	0.2650495

```
elwe$wadr_dummy  
68.5061220
```

(29) 回帰係数をファイルに書き込み (関数 `write.table()`)

```
sep (separator : sep = "," カンマ区切り)
```

```
> write.table(coef(lm_re1), "C:/data/lm_coef1.csv", sep = ",")
```