

Exercise

Code ▼

プロ野球選手の成績を **偏差値** を使って比較

1. data を読み込む
- csv 自体でも内容を確認
2. histograme で確認
- 色々階級を変えて確認
3. 偏差値を求める
- scale を使って標準化

csv file 読み込み

- working dirctory 確認

Hide

```
getwd()
```

Hide

```
[1] "/home/rstudio/r_rstudio/statistics/01_descriptive_statistis"
```

```
data <- read_csv("data01_mac.csv", locale = locale(encoding = "utf8"))
```

Column specification

```
cols(
  .default = col_double(),
  選手名 = col_character(),
  チーム = col_character()
)
i Use `spec()` for the full column specifications.
```

Hide

```
data
```

選手名 <chr>	チーム <chr>	打率 <dbl>	試合 <dbl>	打席数 <dbl>	打数 <dbl>	安打 <dbl>	本塁打 <dbl>	打点 <dbl>	盗塁 <dbl>
秋山 翔吾	西武	0.331	121	560	478	158	21	77	13
マギー	巨人	0.317	117	489	438	139	14	64	4
宮﹂ 敏郎	DeNA	0.315	107	437	403	127	11	52	0
大島 洋平	中日	0.313	119	521	476	149	3	29	23
柳田 悠岐	ソフトバンク	0.311	119	504	405	126	29	92	14
丸 佳浩	広島	0.310	125	571	487	151	21	84	12
安部 友裕	広島	0.306	111	403	369	113	3	38	17
浅村 栄斗	西武	0.305	121	538	486	148	16	88	5
坂本 勇人	巨人	0.302	120	516	453	137	14	57	13
口ペス	DeNA	0.301	120	514	482	145	27	92	0

1-10 of 51 rows | 1-10 of 20 columns

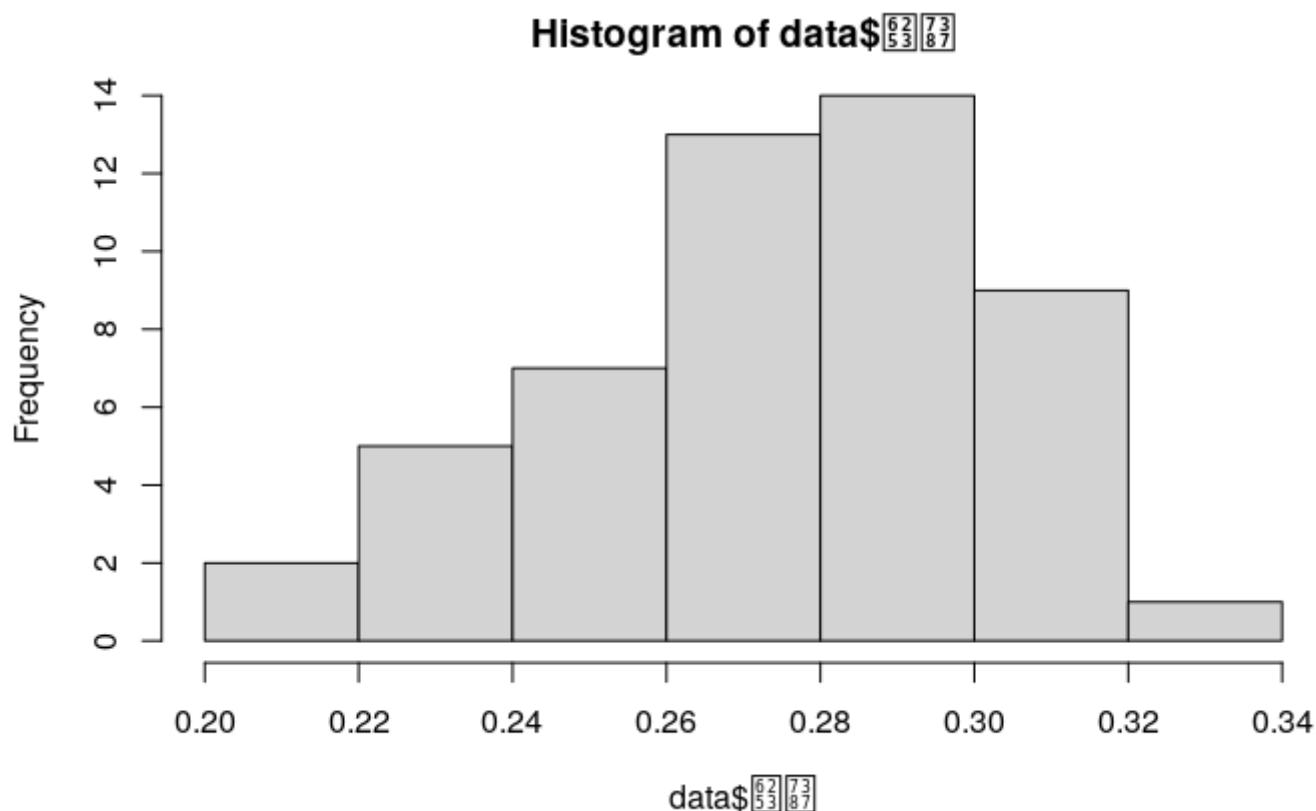
Previous123456Next

histograme で確認

- まずは histograme で確認

Hide

```
hist(data$打率)
```



階級の幅を変えて表示する

- 色々印象が変わってくる

Hide

```
hist(data$打率, breaks = seq(0.2, 0.34, 0.01))
```

偏差値を確認

- 標準化
 - data の 平均値を 0 , 分散を 1 に変換する操作 = 標準化して値の **scale** を整える

$$\frac{X - \mu}{\sigma}$$

$$\text{標準化} = \frac{\text{値} - \text{平均値}}{\text{標準偏差}}$$

標準 : 0 標準偏差 : 1 になる

Hide

```
scale(data$打率)
```

```
[,1]
[1,] 1.967601568
[2,] 1.473451895
[3,] 1.402859085
[4,] 1.332266275
[5,] 1.261673464
[6,] 1.226377059
[7,] 1.085191438
[8,] 1.049895033
[9,] 0.944005817
[10,] 0.908709412
[11,] 0.873413007
[12,] 0.802820197
[13,] 0.767523791
[14,] 0.767523791
[15,] 0.732227386
[16,] 0.732227386
[17,] 0.661634576
[18,] 0.626338171
[19,] 0.626338171
[20,] 0.485152550
[21,] 0.343966929
[22,] 0.343966929
[23,] 0.343966929
[24,] 0.273374119
[25,] 0.061595688
[26,] -0.008997123
[27,] -0.150182744
[28,] -0.150182744
[29,] -0.220775554
[30,] -0.291368364
[31,] -0.291368364
[32,] -0.291368364
[33,] -0.361961175
[34,] -0.361961175
[35,] -0.467850390
[36,] -0.503146796
[37,] -0.503146796
[38,] -0.573739606
[39,] -0.679628822
[40,] -0.679628822
[41,] -0.820814442
[42,] -0.891407253
[43,] -0.926703658
[44,] -1.032592874
[45,] -1.244371305
[46,] -1.279667710
[47,] -1.314964115
[48,] -1.632631762
[49,] -1.703224572
[50,] -2.303263461
[51,] -2.409152676
attr("scaled:center")
[1] 0.2752549
attr("scaled:scale")
[1] 0.0283315
```

偏差値を出す

- 打率の偏差値 = 標準化の値 × 10 + 50

Hide

```
scale(data$打率) * 10 + 50
```

```
[,1]
[1,] 69.67602
[2,] 64.73452
[3,] 64.02859
[4,] 63.32266
[5,] 62.61673
[6,] 62.26377
[7,] 60.85191
[8,] 60.49895
[9,] 59.44006
[10,] 59.08709
[11,] 58.73413
[12,] 58.02820
[13,] 57.67524
[14,] 57.67524
[15,] 57.32227
[16,] 57.32227
[17,] 56.61635
[18,] 56.26338
[19,] 56.26338
[20,] 54.85153
[21,] 53.43967
[22,] 53.43967
[23,] 53.43967
[24,] 52.73374
[25,] 50.61596
[26,] 49.91003
[27,] 48.49817
[28,] 48.49817
[29,] 47.79224
[30,] 47.08632
[31,] 47.08632
[32,] 47.08632
[33,] 46.38039
[34,] 46.38039
[35,] 45.32150
[36,] 44.96853
[37,] 44.96853
[38,] 44.26260
[39,] 43.20371
[40,] 43.20371
[41,] 41.79186
[42,] 41.08593
[43,] 40.73296
[44,] 39.67407
[45,] 37.55629
[46,] 37.20332
[47,] 36.85036
[48,] 33.67368
[49,] 32.96775
[50,] 26.96737
[51,] 25.90847
attr("scaled:center")
[1] 0.2752549
attr("scaled:scale")
[1] 0.0283315
```

それぞれの偏差値を算出

Hide

```
array1 = matrix(0, 51, 5)
array1
```

```
      [,1][,2][,3][,4][,5]
[1,]  0  0  0  0  0
[2,]  0  0  0  0  0
[3,]  0  0  0  0  0
[4,]  0  0  0  0  0
[5,]  0  0  0  0  0
[6,]  0  0  0  0  0
[7,]  0  0  0  0  0
[8,]  0  0  0  0  0
[9,]  0  0  0  0  0
[10,]  0  0  0  0  0
[11,]  0  0  0  0  0
[12,]  0  0  0  0  0
[13,]  0  0  0  0  0
[14,]  0  0  0  0  0
[15,]  0  0  0  0  0
[16,]  0  0  0  0  0
[17,]  0  0  0  0  0
[18,]  0  0  0  0  0
[19,]  0  0  0  0  0
[20,]  0  0  0  0  0
[21,]  0  0  0  0  0
[22,]  0  0  0  0  0
[23,]  0  0  0  0  0
[24,]  0  0  0  0  0
[25,]  0  0  0  0  0
[26,]  0  0  0  0  0
[27,]  0  0  0  0  0
[28,]  0  0  0  0  0
[29,]  0  0  0  0  0
[30,]  0  0  0  0  0
[31,]  0  0  0  0  0
[32,]  0  0  0  0  0
[33,]  0  0  0  0  0
[34,]  0  0  0  0  0
[35,]  0  0  0  0  0
[36,]  0  0  0  0  0
[37,]  0  0  0  0  0
[38,]  0  0  0  0  0
[39,]  0  0  0  0  0
[40,]  0  0  0  0  0
[41,]  0  0  0  0  0
[42,]  0  0  0  0  0
[43,]  0  0  0  0  0
[44,]  0  0  0  0  0
[45,]  0  0  0  0  0
[46,]  0  0  0  0  0
[47,]  0  0  0  0  0
[48,]  0  0  0  0  0
[49,]  0  0  0  0  0
[50,]  0  0  0  0  0
[51,]  0  0  0  0  0
```

計算結果を入れていく

- columns に代入 **[行, 列]**

Hide

```
array1[,1] = scale(data$打率) * 10 + 50
array1[,2] = scale(data$本塁打) * 10 + 50
array1[,3] = scale(data$打点) * 10 + 50
array1[,4] = scale(data$盗塁) * 10 + 50
array1[,5] = scale(data$出塁率) * 10 + 50

array1
```

```

    [,1] [,2] [,3] [,4] [,5]
[1,] 69.67602 57.12759 61.32293 55.94369 69.02193
[2,] 64.73452 49.41683 54.53531 45.72588 58.69542
[3,] 64.02859 46.11222 48.26982 41.18464 55.25325
[4,] 63.32266 37.29992 36.26097 67.29681 54.39270
[5,] 62.61673 65.93989 69.15479 57.07900 73.89834
[6,] 62.26377 57.12759 64.97779 54.80838 64.14552
[7,] 60.85191 37.29992 40.96009 60.48494 49.51630
[8,] 60.49895 51.61991 67.06629 46.86120 52.67162
[9,] 59.44006 49.41683 50.88044 55.94369 59.84281
[10,] 59.08709 63.73682 69.15479 41.18464 44.06619
[11,] 58.73413 62.63528 68.11054 59.34963 61.27705
[12,] 58.02820 38.40146 40.96009 49.13182 63.28498
[13,] 57.67524 37.29992 37.82734 45.72588 58.40857
[14,] 57.67524 60.43221 57.14593 44.59057 54.96640
[15,] 57.32227 41.70607 37.82734 79.78524 59.26911
[16,] 57.32227 50.51837 42.00434 44.59057 58.40857
[17,] 56.61635 39.50300 48.26982 74.10868 61.27705
[18,] 56.26338 37.29992 43.57071 42.31995 55.25325
[19,] 56.26338 43.90915 46.18133 41.18464 52.38477
[20,] 54.85153 58.22913 63.93355 42.31995 66.15346
[21,] 53.43967 48.31530 45.13708 51.40244 50.37684
[22,] 53.43967 63.73682 51.40257 41.18464 63.57183
[23,] 53.43967 47.21376 47.74770 58.21432 60.12966
[24,] 52.73374 40.60454 42.00434 41.18464 42.91880
[25,] 50.61596 70.34604 63.93355 41.18464 45.21358
[26,] 49.91003 48.31530 48.79195 49.13182 38.90294
[27,] 48.49817 37.29992 36.78309 44.59057 48.94260
[28,] 48.49817 45.01068 46.18133 45.72588 52.09792
[29,] 47.79224 65.93989 57.14593 41.18464 52.95847
[30,] 47.08632 46.11222 46.70345 54.80838 41.19772
[31,] 47.08632 38.40146 37.30522 65.02619 36.32131
[32,] 47.08632 50.51837 57.66806 41.18464 45.21358
[33,] 46.38039 62.63528 56.62381 49.13182 45.21358
[34,] 46.38039 49.41683 40.43796 47.99651 46.93467
[35,] 45.32150 47.21376 42.00434 43.45526 48.08206
[36,] 44.96853 36.19839 41.48221 44.59057 34.31338
[37,] 44.96853 51.61991 54.53531 42.31995 53.53216
[38,] 44.26260 37.29992 45.65920 76.37931 39.47663
[39,] 43.20371 65.93989 64.45567 44.59057 46.93467
[40,] 43.20371 41.70607 37.82734 58.21432 46.07413
[41,] 41.79186 58.22913 52.96894 46.86120 37.46870
[42,] 41.08593 45.01068 41.48221 62.75556 39.18978
[43,] 40.73296 53.82298 48.79195 59.34963 43.77934
[44,] 39.67407 53.82298 48.26982 42.31995 38.32924
[45,] 37.55629 53.82298 48.26982 41.18464 39.47663
[46,] 37.20332 36.19839 36.78309 43.45526 40.05033
[47,] 36.85036 57.12759 54.53531 55.94369 53.24531
[48,] 33.67368 65.93989 62.88930 41.18464 38.61609
[49,] 32.96775 60.43221 57.66806 41.18464 42.34511
[50,] 26.96737 51.61991 53.49106 41.18464 35.46077
[51,] 25.90847 35.09685 32.60610 43.45526 31.44490
```

各選手の情報を取得

- index で行を指定して取得

Hide

```
round(array1[1,])
```

```
[1] 70 57 61 56 69
```

- 偏差値に落とし込むことによって各種を比較できる

打率	本塁打	打点	盗塁	出塁率
70	57	61	56	69