

Exercise3

Code ▼

問題

- お店の評価を分析
 - アンケートの総合評価から **cs分析** と **重回帰分析** を行い,改善点と総合評価に関する要因を見つける

満足度アンケート

設問	評価
1. 料理の味	1~5
2. menu種類	1~5
3. 清潔感	1~5
4. 雰囲気	1~5
5. 価格	1~5
6. 対応	1~5
7. 総合評価	1~5

解析手順

1. data を読み込む
 - **csv 自体でも内容を確認**
2. 散布図と相関係数を確認
 - **plot**で散布図, **cor**で相関係数を求める
3. CS分析
 - **各要素の平均値と相関係数を比較**
4. 重回帰分析
 - **lm** を使って総合評価に与える影響を求める

1. data を読み込む

- **csv 自体でも内容を確認**

Hide

```
getwd()
```

```
[1] "/home/rstudio/r_rstudio/statistics/03_correlation_regression"
```

Hide

```
data = read.csv("data03_mac.csv")
data
```

設問1 <int>	設問2 <int>	設問3 <int>	設問4 <int>	設問5 <int>	設問6 <int>	総合評価 <int>
4	4	1	3	5	1	4
3	4	5	1	4	2	3
5	2	5	1	5	2	5
5	2	3	3	4	2	5

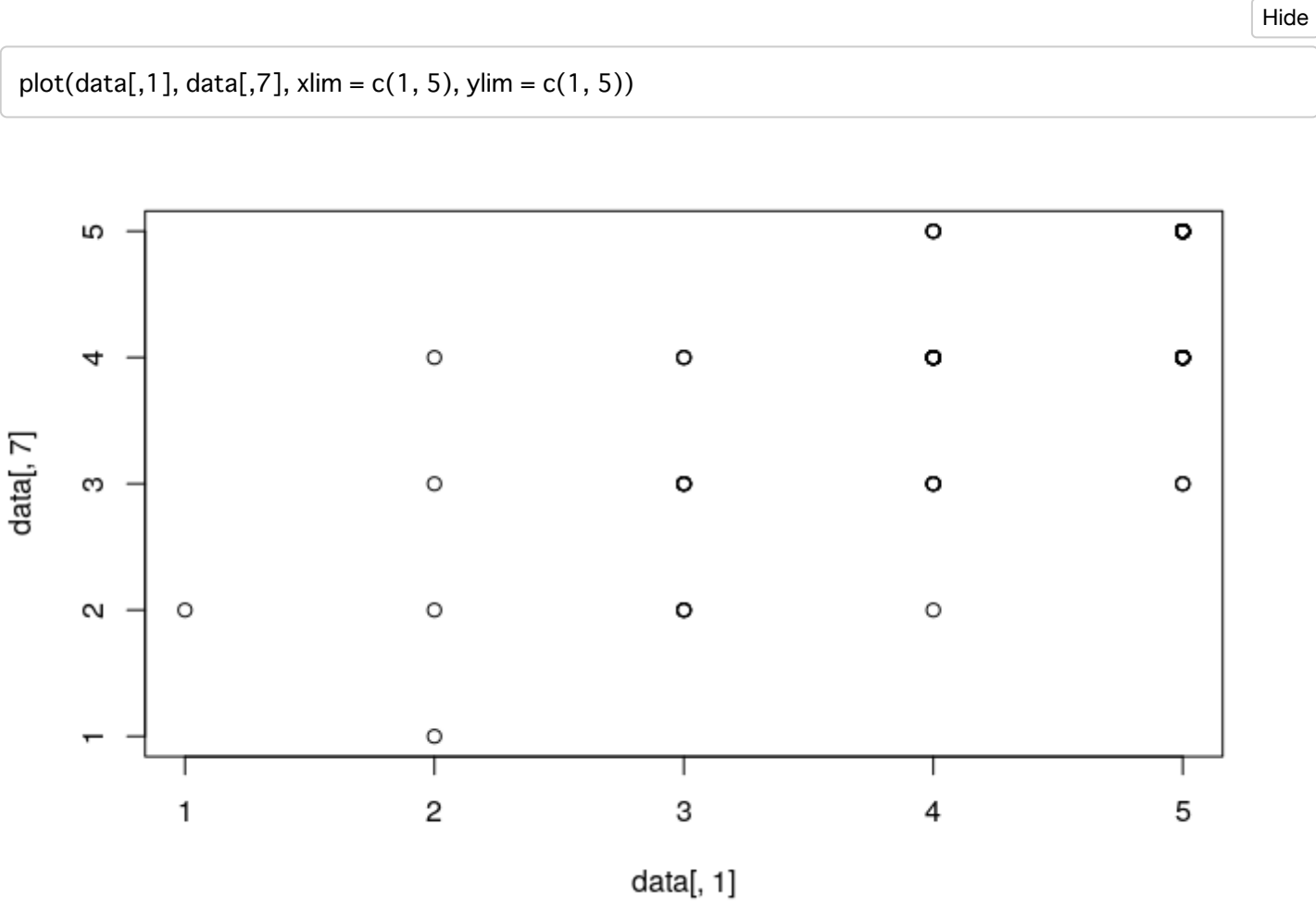
設問1 <int>	設問2 <int>	設問3 <int>	設問4 <int>	設問5 <int>	設問6 <int>	総合評価 <int>
5	1	5	1	4	2	5
5	3	2	1	5	5	5
5	1	5	5	3	2	4
5	1	1	5	3	1	4
5	2	5	1	4	3	5
4	3	5	1	4	1	3

1-10 of 100 rows

Previous123456...10Next

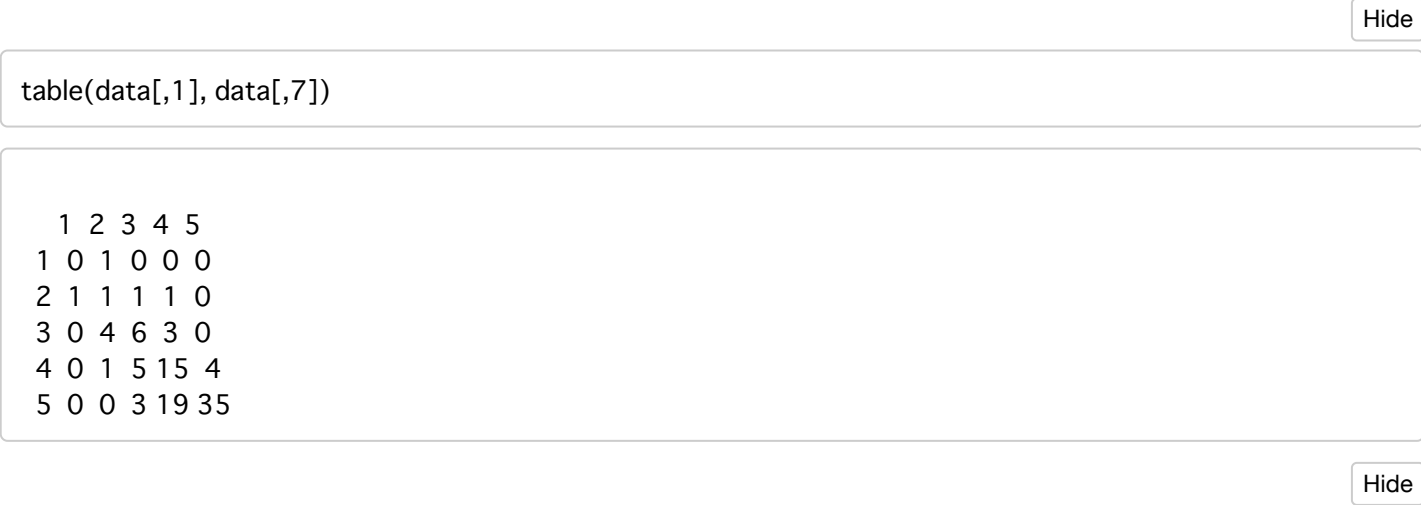
散布図 plot

- x軸：設問1 の結果
- y軸：総合評価



離散データ用に plot

- 重なっている所の数を表示してくれる
 - クロス集計表を使用
- 相関係数 : 0.7070379



```
cor(data[,1], data[,7])
```

```
[1] 0.7070379
```

設問2 の結果

Hide

```
rnum <- 2
table(data[,rnum], data[,7])
```

```
  1 2 3 4 5
1 1 4 6 14 10
2 0 2 3 9 11
3 0 1 4 7 5
4 0 0 2 6 7
5 0 0 0 2 6
```

Hide

```
cor(data[,rnum], data[,7])
```

```
[1] 0.2530099
```

設問3 の結果

Hide

```
rnum <- 3
table(data[,rnum], data[,7])
```

```
  1 2 3 4 5
1 0 0 2 3 1
2 1 0 3 5 3
3 0 1 2 0 4
4 0 4 0 8 8
5 0 2 8 22 23
```

Hide

```
cor(data[,rnum], data[,7])
```

```
[1] 0.1608865
```

設問4 の結果

Hide

```
rnum <- 4
table(data[,rnum], data[,7])
```

```
  1 2 3 4 5
1 1 5 5 18 23
2 0 0 2 4 4
3 0 0 2 6 5
4 0 1 2 0 2
5 0 1 4 10 5
```

Hide

```
cor(data[,rnum], data[,7])
```

```
[1] -0.0734688
```

設問5の結果

Hide

```
rnum <- 5
table(data[,rnum], data[,7])
```

```
  1 2 3 4 5
2 1 1 2 0 0
3 0 3 7 9 0
4 0 3 6 23 19
5 0 0 0 6 20
```

Hide

```
cor(data[,rnum], data[,7])
```

```
[1] 0.6335291
```

設問6の結果

Hide

```
rnum <- 6
table(data[,rnum], data[,7])
```

```
  1 2 3 4 5
1 1 6 5 12 1
2 0 1 10 22 11
3 0 0 0 4 19
4 0 0 0 0 7
5 0 0 0 0 1
```

Hide

```
cor(data[,rnum], data[,7])
```

```
[1] 0.623389
```

3. CS分析

- 各要素の平均値と相関係数を比較
 - CS分析の場合はまずは **相関係数** と **平均値** を求め → **散布図**

各相関係数と平均を出力

- 相関係数

Hide

```
x3 <- cor(data[, 1:6], data[, 7])
x3
```

```
[,1]
設問1 0.7070379
設問2 0.2530099
設問3 0.1608865
設問4 -0.0734688
設問5 0.6335291
設問6 0.6233890
```

- 平均値

Hide

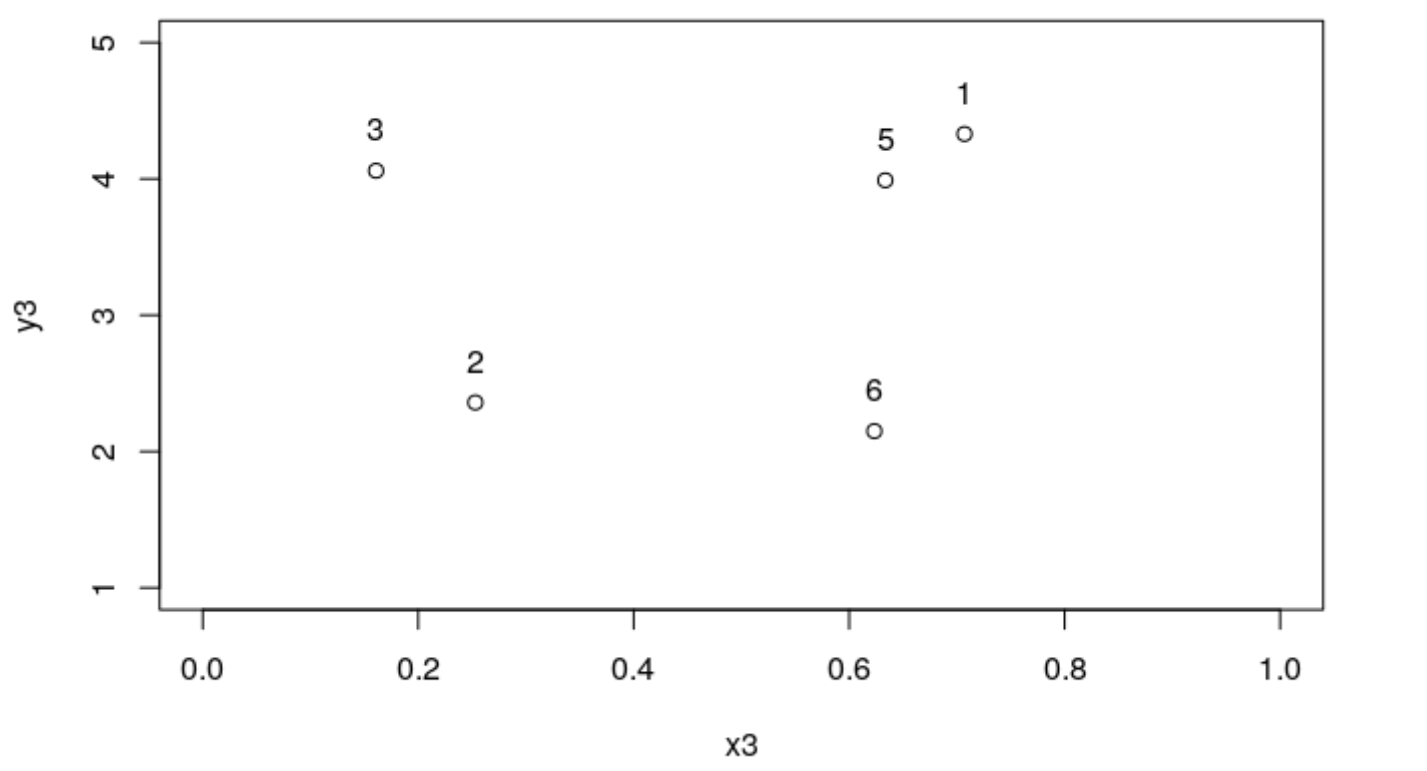
```
y3 <- colMeans(data[, 1:6])
y3
```

```
設問1 設問2 設問3 設問4 設問5 設問6
4.33 2.36 4.06 2.31 3.99 2.15
```

- 散布図

Hide

```
plot(x3, y3, xlim = c(0, 1), ylim = c(1, 5))
text(x3, y3 +0.3, labels = c(1:6))
```



- 相関係数 : x3
- 平均値 : y3
- 1 ~ 6 : 各設問 1 ~ 6

平均値高い + 相関係数も高い

||

とても重要な項目である

- 1 (料理の味) と 5 (価格) はこのお店にとって 重要性大

平均値低い + 相関係数も高い

||

この項目のせいで総合評価が下がっている

||

この項目をまず改善しないといけない

- まず改善すべきは, お客様に対するスタッフのサービス(設問6)

設問	評価
1. 料理の味	1~5
2. menu種類	1~5
3. 清潔感	1~5
4. 雰囲気	1~5
5. 価格	1~5
6. 対応	1~5
7. 総合評価	1~5

相関係数 と 平均値 を組み合わせる事によって

↓

改善点 と クオリティーの保持する所（力を入れるべき所）がわかる

4. 重回帰分析

- 必ず全ての項目の相関係数を相互に確認をする

Hide

round(cor(data), 2)

	設問1	設問2	設問3	設問4	設問5	設問6	総合評価
設問1	1.00	0.13	0.12	0.02	0.41	0.43	0.71
設問2	0.13	1.00	0.10	-0.02	0.27	0.19	0.25
設問3	0.12	0.10	1.00	-0.26	0.15	0.10	0.16
設問4	0.02	-0.02	-0.26	1.00	-0.22	-0.14	-0.07
設問5	0.41	0.27	0.15	-0.22	1.00	0.45	0.63
設問6	0.43	0.19	0.10	-0.14	0.45	1.00	0.62
総合評価	0.71	0.25	0.16	-0.07	0.63	0.62	1.00

- 設問1, 設問5, 設問6, この3つに対してはそこそこ高い 相関係数 があるので
 - 3つで重回帰分析を行う
- 設問同士 で高い相関係数のものが無いかも確認!!
 - 相関係数が高い = 同じ様な category である

重回帰分析を実装

Hide

res3 <- lm(data[,7]~data[,1]+data[,5]+data[,6])
summary(res3)

Call:
lm(formula = data[, 7] ~ data[, 1] + data[, 5] + data[, 6])

Residuals:

Min	1Q	Median	3Q	Max
-1.87416	-0.32555	0.03175	0.35942	1.12584

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.15879	0.31989	-0.496	0.621
data[, 1]	0.47098	0.06674	7.057	2.65e-10 ***
data[, 5]	0.38953	0.07926	4.914	3.66e-06 ***
data[, 6]	0.29544	0.06884	4.292	4.23e-05 ***

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

Residual standard error: 0.5321 on 96 degrees of freedom
Multiple R-squared: 0.6997, Adjusted R-squared: 0.6903
F-statistic: 74.56 on 3 and 96 DF, p-value: < 2.2e-16

- **Multiple R-squared: 0.6997** : 当てはまりが良いのどうかの指標!!
 - 約 0.7 くらいの当てはまりなので, そこそこ良い重回帰分析ができている
- 係数

Hide

res3

Call:
lm(formula = data[, 7] ~ data[, 1] + data[, 5] + data[, 6])

Coefficients:

(Intercept)	data[, 1]	data[, 5]	data[, 6]
-0.1588	0.4710	0.3895	0.2954