

心理・教育測定法 基礎編 B 多変量解析入門

回帰分析

バイプロット

2014年度 心理・教育測定法 A・B、統計演習 A・B

心理・教育測定法(基礎編) A・B
金曜日 5・6限:W9-935

人間行動システム統計演習A・B
木曜日 5・6限 W9-607

心理・教育測定法(基礎編) A・B		担当	人間行動システム統計演習A・B		担当
4月11日	オリエンテーション		4月10日	オリエンテーション	
18日	統計量の計算・相関と回帰		17日	データフレームの操作等	
25日	統計的推測の基礎		24日	ヒストグラム、クロス集計等	
5月2日	平均値に関する推測(1)	心理・教育測定法A	5月15日	相関と回帰	統計演習A
9日	平均値に関する推測(2)	前川	22日	平均値に関する推測	中川、寺井、和嶋
23日	分散分析(1)		29日	分散分析	
30日	分散分析(2)		6月5日	分散分析	
6月6日	イントロ		6月12日	イントロ	
13日	多変量データの取り扱い		19日	多変量グラフ	
20日	回帰分析	心理・教育測定法B	26日	回帰分析	統計演習B
27日	クラスター分析	前川	7月3日	クラスター分析	前川、栗山、荒井
7月4日	多次元尺度法		10日	多次元尺度法	
11日	因子分析(1)		17日	因子分析等	
18日	因子分析(2)		24日	因子分析等、Amos, R	

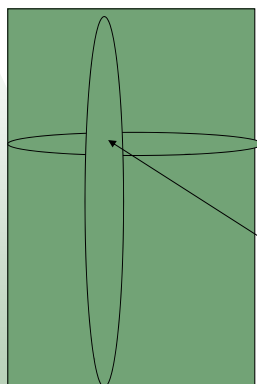
人間行動システム統計演習A・Bには、ノートPCをご持参下さい

多変量データ

j 列 $j=1,2,\dots,p$

i行

$i=1, 2, 3, \dots, n$



y_{ij}

個体 i の変数 j による測定値

バイプロット

$Y = X\beta + \varepsilon$ という線形モデルで説明変数行列の列がふたつの場合。

(切片は数えない。)

バイプロットを作成して、データを視覚化することが出来る。

バイプロット(1)

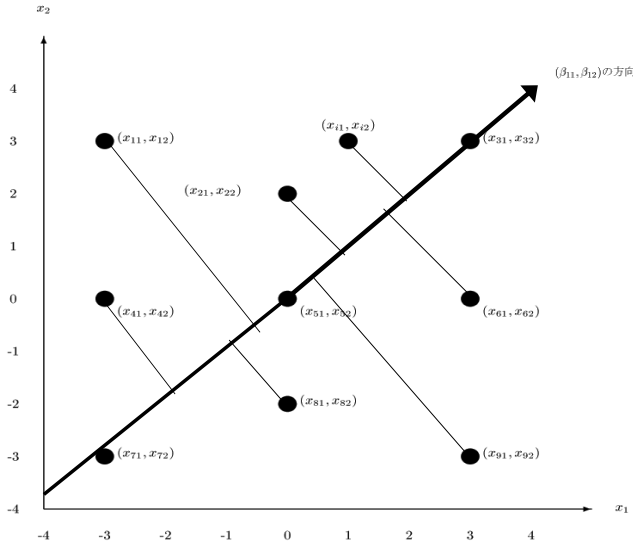


図 1: 線形モデルの図的解釈 (バイプロット・同時プロット)

3
i
6
2
9
5
1
8
4
7

バイプロット(2)

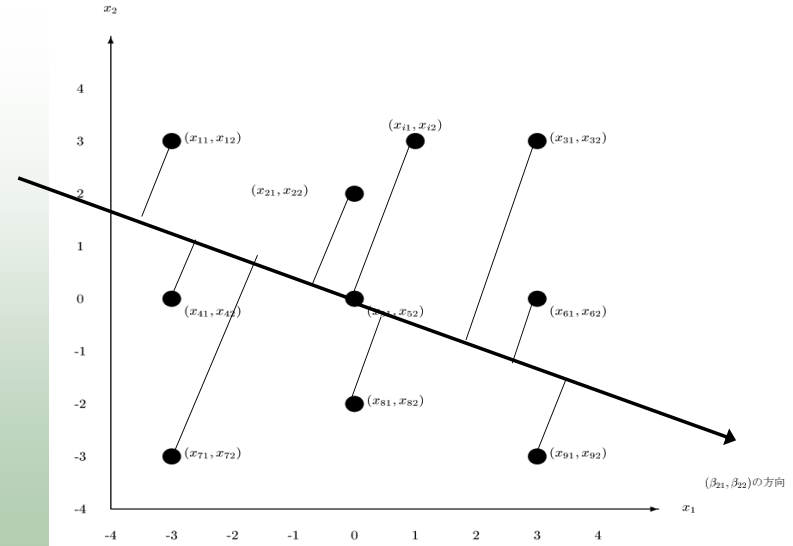


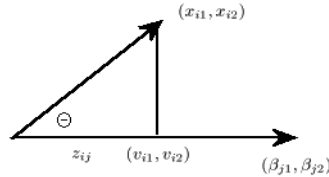
図 1: 線形モデルの図的解釈 (バイプロット・同時プロット)

9
6
3
8
5
i
2
7
4
1

バイプロットと回帰モデル

個体 (x_{i1}, x_{i2}) を点で、変数 (β_{j1}, β_{j2}) を原点からのベクトルで表す。点 (x_{i1}, x_{i2}) からベクトルへ下ろした垂線の足を (v_{i1}, v_{i2}) とし、原点からその垂線の足までの長さを z_{ij} とする。ベクトル β_j と原点から点 (x_{i1}, x_{i2}) に至るベクトルのなす角度を θ とすれば、 $\cos \theta$ とベクトルの内積の関係から

$$\begin{aligned} z_{ij} &= \|x_i\| \cos \theta \\ &= \|x_i\| \frac{x_i' \beta_j}{\|x_i\| \|\beta_j\|} \\ &= \frac{x_i' \beta_j}{\|\beta_j\|} \end{aligned}$$



となる。
従って、 \hat{y}_{ij} が z_{ij} に比例するというモデル

$$\hat{y}_{ij} = a + b z_{ij}$$

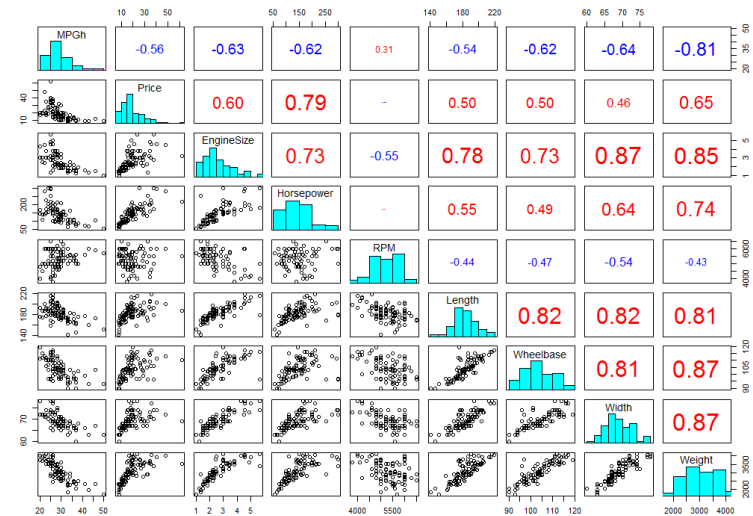
を考えると

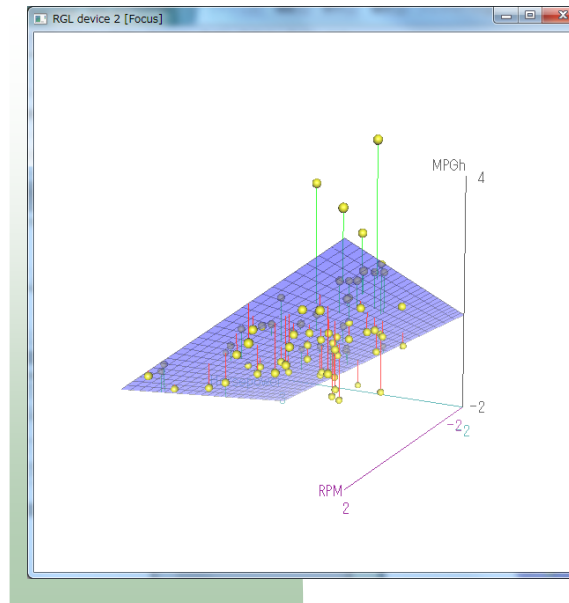
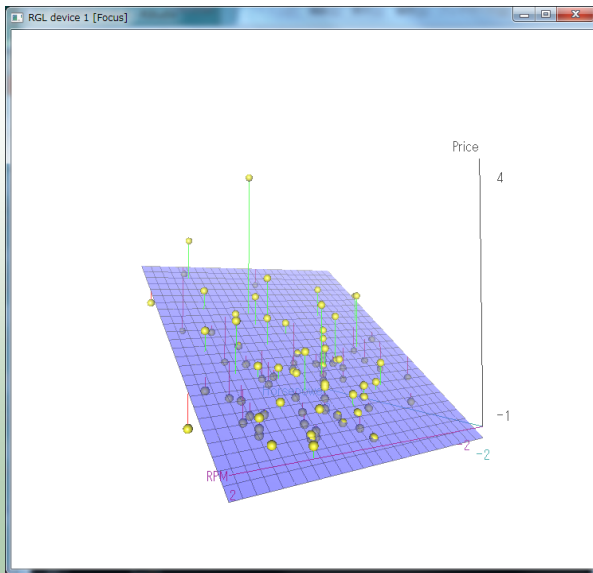
$$\begin{aligned} \hat{y}_{ij} &= a + b \frac{x_i' \beta_j}{\|\beta_j\|} \\ &= a + x_i' \beta_j' \end{aligned}$$

と回帰分析の形に書くことができる。

従って、データを因子分析し、
因子スコアと点、因子パターンをベクトルとして
プロットすればよい。

散布図行列 + α





```
res1=lm(MPGh~RPM+Horsepower)$coefficients
res2=lm(Price~RPM+Horsepower)$coefficients
res3=lm(EngineSize~RPM+Horsepower)$coefficients
res4=lm(Weight~RPM+Horsepower)$coefficients
```

```
coef=rbind(res1, res2, res3, res4)
rownames(coef)=c("MPGh", "Price", "EngineSize", "Weight")
coef=coef[, 2:3]
```

	RPM	Horsepower
MPGh	0.33663345	-0.6313942
Price	-0.03391888	0.7894620
EngineSize	-0.57553265	0.7532350
Weight	-0.45564995	0.7555145

```
# native plot
plot(RPM, Horsepower)
text(RPM, Horsepower, rownames(cars93n))
arrows(rep(0, 4), rep(0, 4), 2*coef[, 1], 2*coef[, 2],
       col=c("red", "green", "blue", "magenta"),
       lw=2)
text(2.1*coef[, 1], 2.1*coef[, 2], c("MPGh", "Price", "EngineSize", "Weight"),
     col=c("red", "green", "blue", "magenta"))
```

```
# rowd cloud
library(wordcloud)
textplot(RPM, Horsepower, rownames(cars93n))
arrows(rep(0, 4), rep(0, 4), 2*coef[, 1], 2*coef[, 2],
       col=c("red", "green", "blue", "magenta"),
       lw=2)
text(2.1*coef[, 1], 2.1*coef[, 2], c("MPGh", "Price", "EngineSize", "Weight"),
     col=c("red", "green", "blue", "magenta"))
```

```
# native biplot
Xmat=cbind(RPM, Horsepower); rownames(Xmat)=rownames(cars93ns)
Ymat=coef
biplot(Xmat, Ymat, col=c("red", "green", "blue", "yellow"))
```

