

# 生体情報工学演習 第15回

北海道大学工学部情報エレクトロニクス学科

生体情報コース3年

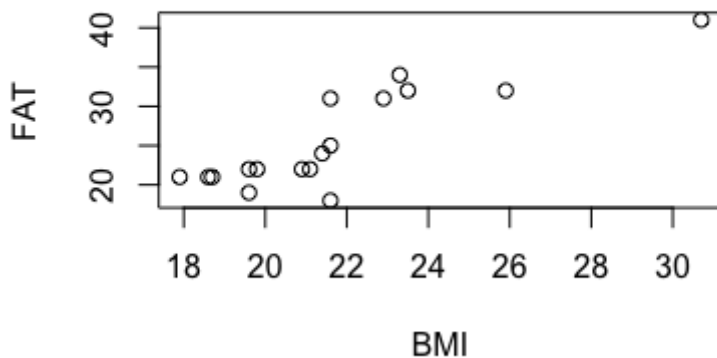
学生番号: 02180144 茂木貴紀

作成日: 2021/01/28

## No1

まず, 以下を実行して, 体脂肪率とBMIの関係を可視化させた.

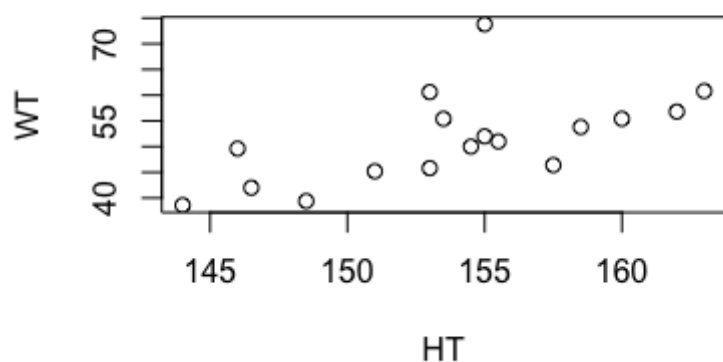
```
> data <- read.delim('R10.txt')  
> attach(data)  
> plot(FAT~BMI)
```



以上から, 正の相関があると考えられる. `cor.test(FAT, BMI)` を実行すると, p値が $6.6e-06$  となり, 帰無仮説が棄却され, 相関係数は0に近くないとわかる. 実際に出力されたピアソンの相関係数の推定値は0.86...となり, 正の相関性が高いと判断できる.

次に, 身長を独立変数 (x), 体重を従属変数 (y) として, 回帰分析を行う. 以下を実行.

```
> plot(WT~HT)  
> res <- lm(WT~HT)  
> summary(res)
```



結果は, p値が0.01で, 有意確率を0.05とすれば帰無仮説が棄却されることとなるものの, 決定係数が0.31 とそこまで高くなく, 正の相関が大きいとまではいえない。

また, 以下のようにして, 身長が155cmの場合の体重を95%信頼区間とともに推定する。

```
> predict(res, list(HT=155), interval="confidence")
```

以上から, 体重は52.6 [48.8:56.5] kgと推定される。

## No2

まず, 以下のようなtableを作成した。

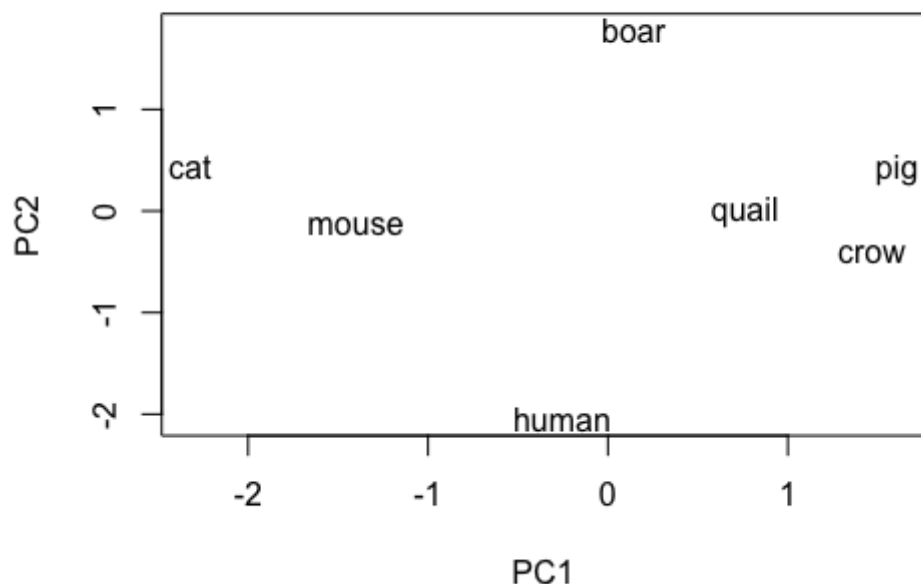
```
> data2 <- data.frame(
  row.names = c('human', 'cat', 'mouse', 'pig', 'boar', 'crow', 'quail'),
  genA= c(1.53, 3.22, 2.94, 2.33, 2.98, 1.96, 2.75),
  genB=c(2.19, 2.87, 3.18, 1.29, 0.97, 3.91, 1.05),
  genC=c(0.54, 1.91, 1.22, 0.31, 1.34, 0.69, 0.12),
  genD = c(3.19, 3.17, 2.98, 2.01, 1.78, 2.08, 2.94),
  genE = c(0.11, 0.49, 0.61, 1.99, 1.27, 2.81, 1.82)
)
```

	genA	genB	genC	genD	genE
human	1.53	2.19	0.54	3.19	0.11
cat	3.22	2.87	1.91	3.17	0.49
mouse	2.94	3.18	1.22	2.98	0.61
pig	2.33	1.29	0.31	2.01	1.99
boar	2.98	0.97	1.34	1.78	1.27
crow	1.96	3.91	0.69	2.08	2.81
quail	2.75	1.05	0.12	2.94	1.82

これを, 主成分分析する.

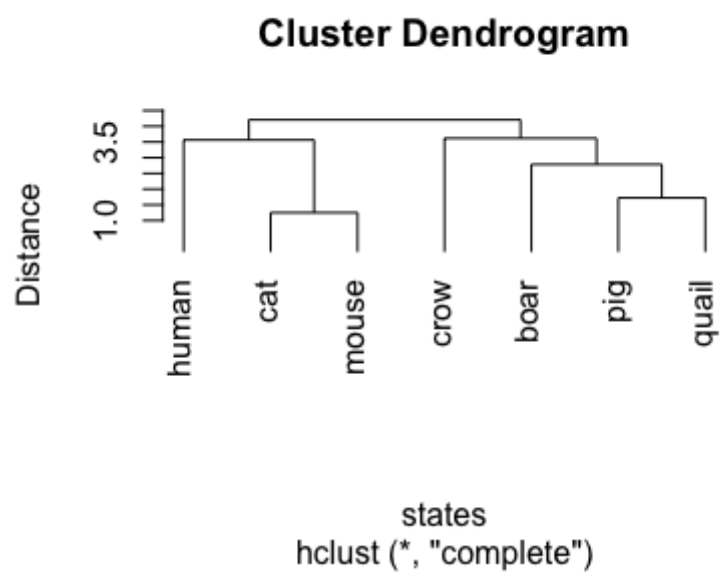
```
> res <- prcomp(data, scale=T)
> summary(res)
> plot(res$x[,1],res$x[,2],type="n",xlab="PC1", ylab="PC2")
> text(res$x[,1],res$x[,2],rownames(res$x))
```

寄与率は, 第一主成分で0.425, 第二種成分で0.691となっている. ちなみに第三種成分では0.91となっているため, 本当は3次元上で可視化するのが適当である.



最後に, これらの遺伝子に関してクラスター分析を行い, 樹形図を作成する. 以下のコマンドを実行する.

```
> states.d <- dist(res$x)
> states.out <- hclust(states.d, method="complete")
> plot(states.out, hang=-1, xlab="states", ylab="Distance")
```



tags: 生体情報工学演習