

生体情報工学演習 第12回

北海道大学情報エレクトロニクス学科

生体情報コース3年

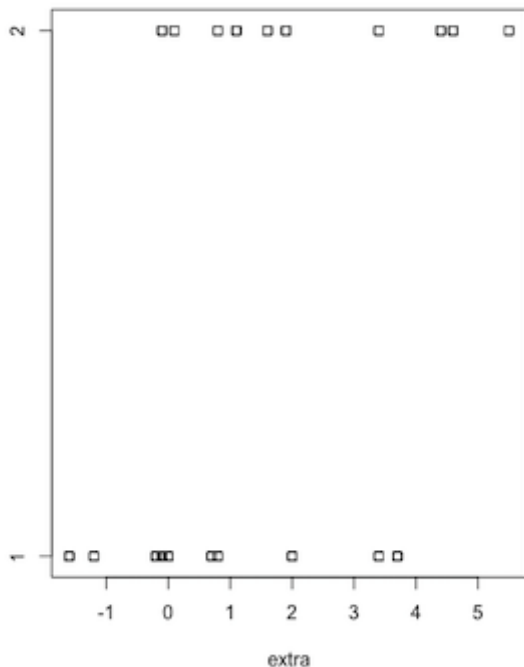
学生番号: 02180144 茂木貴紀

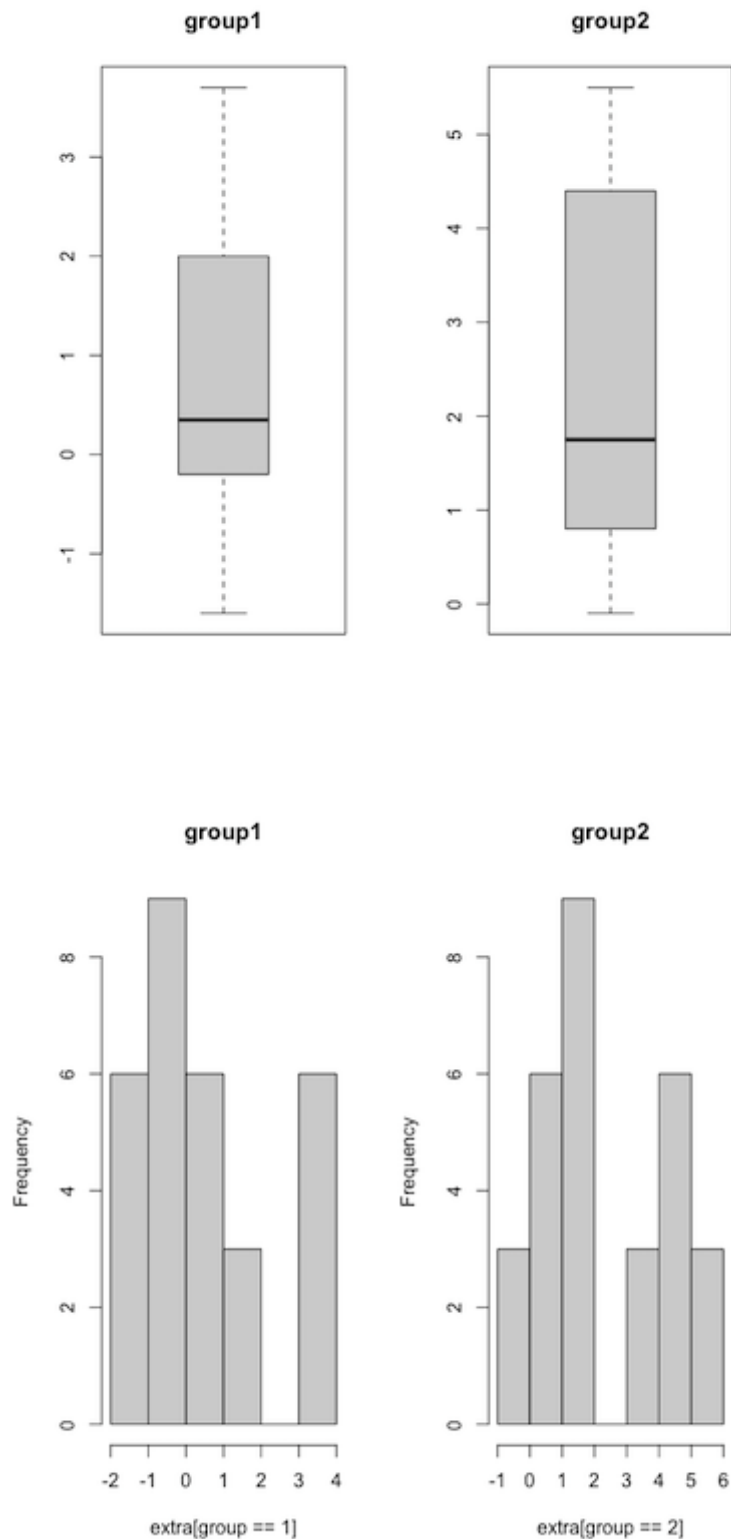
作成日: 2020/12/25

課題1

まず, 分布を知るために, 1次元散布図, ヒストグラム, 箱ひげ図を作成した. コード及びグラフは以下の通りである.

```
> dat <- read.delim('./R05.txt')
> attach(dat)
> stripchart(extra~group)
> par(mfrow=c(1,2))
> boxplot(extra[group==1], main="group1")
> boxplot(extra[group==2], main="group2")
> hist(extra[group==1], main="group1")
> hist(extra[group==2], main="group2")
```





まず, 散布図から, group2の方が催眠時間が長いことが視覚的にわかる. ただ, 散布図を参照したのちに箱ひげ図を参照しても, group2の方が分散が多いことは断定しづらい. また, ヒストグラムを参照しても, お互いのグラフの形状にあまり違いはなく, 特徴が読み取れないことがわかる.

そこで, 平均と標準偏差から参照する. 中央値と四分位範囲を求めなかったのは, それが全ての値を正しく反映できていないからである. (データをソートし, そこから〇番目などの順位からしか判断されず, それ以外のデータは影響されていない)

コードは以下の通りである。

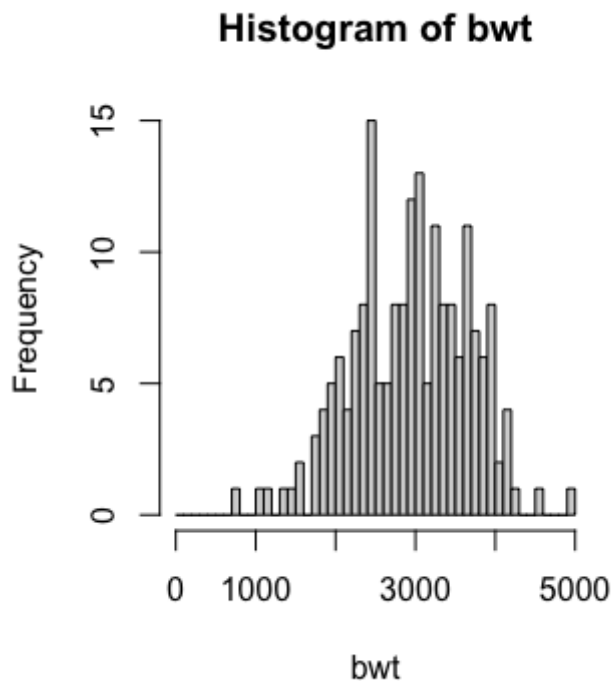
```
> mean(extra[group==1])  
// 0.75  
> mean(extra[group==2])  
// 2.33  
> sd(extra[group==1])  
// 1.72  
> sd(extra[group==2])  
// 1.93
```

以上から、標準偏差はほぼ一致しており、分散がほぼ異なっていないことがわかる。つまり、group1と2で異なるのは全体的な催眠時間である。

課題2

まず、以下を実行する。

```
> library(MASS)  
> attach(birthwt)  
> hist(bwt, breaks=seq(0,5000,100))
```



これが正規分布に従っているかを調査する. ヒストグラムの形状的には従っていそうなので, 以下を実行してp値を取得する.これは, shapiro-wilkの検定と呼ばれ, その集団が正規性を持つか判断するのに役立つ.

```
> shapiro.test(bwt)
// W = 0.99244, p-value = 0.4353
```

p値が0.05を大きく上回っているので, 帰無仮説は棄却されず, 正規分布としてみなして問題ないと言える.

次に, 信頼区間を推定する.

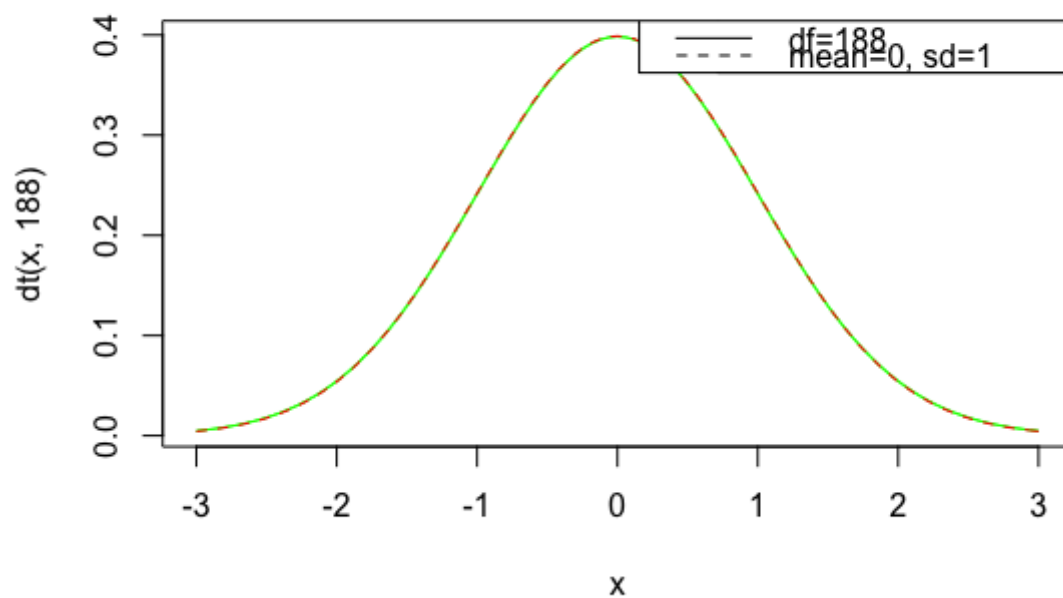
```
> t.test(bwt)
```

ここから, t値が55.514と求まり, サンプル数と不偏標準偏差から母平均の信頼区間を推定することができる (今回は上記のコマンドを入力することで, すぐに得られるが) .

信頼区間は, 2839.952から3049.222の間となる.

さらに, このサンプルは自由度が188と高く, t分布を作成すると, 標準正規分布に非常に近くなっていることがわかる.

```
> curve(dt(x,188),-3,3, add=TRUE, col="green")
> curve(dnorm(x,0,1), -3,3, add=TRUE, col="red", lty=2)
> legend("topright",lty=1:2,legend=c('df=188','mean=0, sd=1'))
```



このグラフの通りである.

tags: 生体情報工学演習