医学情報処理演習第5回「データの分布と検定の概念」2007年11月5日

前回の課題の回答例

解答に最低限必要なプログラムは次の通り。

```
library(MASS)
t.test(Melanoma$age)
```

次の枠内の結果が得られるので,平均年齢は 52.5 歳,その 95% 信頼区間は [50.1 歳,54.8 歳] といえる。有効数字より一桁多く表示する際の四捨五入・切上げ・切捨てのどれを適用すべきかについては,通常,四捨五入でよいのだが,95% 信頼区間は広めに(というか,少なくとも真の 95% 信頼区間を含むように)表示するのが普通なので,ここでは下限は切り捨て,上限は切り上げる。

```
One Sample t-test

data: Melanoma$age

t = 45.0561, df = 204, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

50.16761 54.75922

sample estimates:
mean of x

52.46341
```

なお,別の解き方として,以下も可能である。

```
library(MASS)
print(bar.age <- mean(Melanoma$age))
sd.age <- sd(Melanoma$age)
t975 <- qt(0.975,length(Melanoma$age)-1)
rootn <- sqrt(length(Melanoma$age))
print(bar.age - t975*sd.age/rootn)
print(bar.age + t975*sd.age/rootn)</pre>
```

実行すると,平均年齢,95% 信頼区間の下限,95% 信頼区間の上限が順に表示される。

本日の課題

MASS ライブラリには,低体重出生についてのデータフレーム birthwt が含まれている。そのうち出生体重を示す変数 bwt が正規分布に従っているといえるかどうか,作図により検討した上で検定せよ。検定の有意水準は 5% とする。図を Word または Powerpoint などに貼り付け,検定結果について考察する文章と学籍番号も打って印刷し,署名して提出すること。