

学籍番号:

公衆衛生学

疫学演習 2019-6-5 & 2019-6-12

氏名:

王 超辰 (<https://wangcc.me>)

1 問題1: 両群間計量データの平均値を比較する

200名の認知症患者を募集し, 認識能力テスト(cognitive test, COG), 及び脳萎縮の進行度 (brain atrophy, 脳体積の平均年間減少率, 単位は%) の検査を全員に行った. COG, 及び脳萎縮のデータは大きいほど認知症の進行度がより進んでいる. また, この200名の参加者から採取した血液検体を利用して, ある遺伝子の変異の有無を検査した. このデータは以下の表でまとめた:

変数	遺伝変異あり (n = 50)		遺伝変異なし (n = 150)	
	平均値 (mean)	標準偏差 (standard deviation)	平均値 (mean)	標準偏差 (standard deviation)
認識能力テスト, COG	69.2	9.0	60.2	9.0
脳萎縮度, atrophy, %/year	0.67	0.21	0.23	0.10

1. 帰無仮説を「遺伝子変異ありと変異なし両群の間に, COGの平均値は等しい」とする. 上記のデータ及び適宜な方法を使って検定せよ. 検定の結果を分かりやすく説明せよ. なお, 分散が等しいと仮定できる場合, 以下の式で両群の共通標準偏差が計算できる:

$$S = \sqrt{\frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2}} \quad (1)$$

- S_A : A群の標準偏差;
- n_A : A群の人数;
- S_B : B群の標準偏差;
- n_B : B群の人数;
- S : A群及びB群の共通標準偏差;
- $n_A + n_B - 2$: 分散が等しい時の自由度.

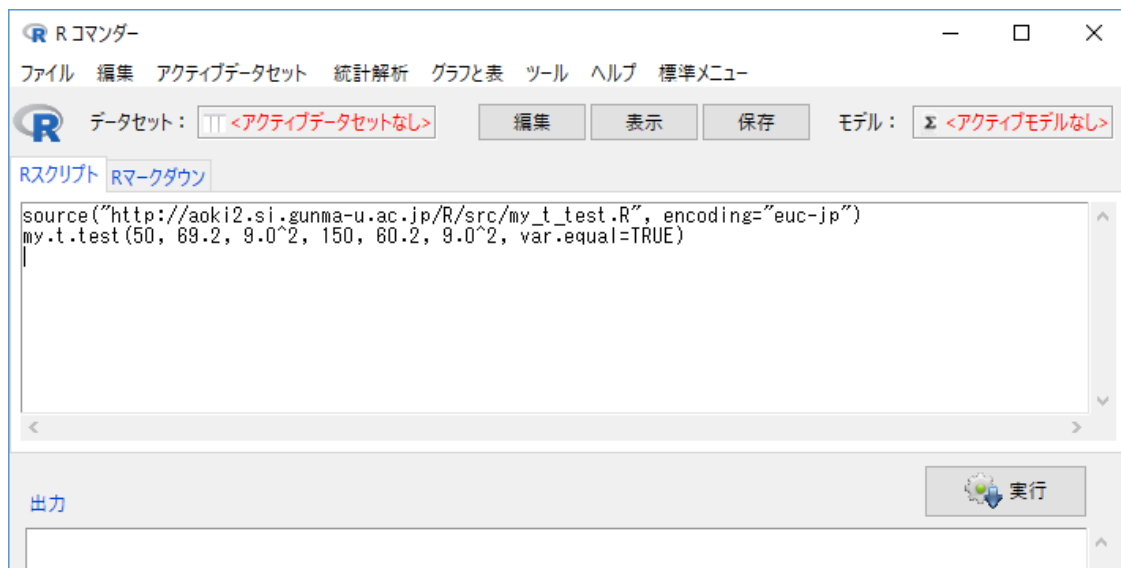
また, EZR で t 値, 自由度 (degree of freedom) を使って P 値を計算する時, 以下のコマンドを利用してください:

```
2*pt(t value, degree of freedom, lower=FALSE)
```

1.1 答え

以下のコードをRスクリプトに入力して,実行をクリックしてください.自分の検定結果とは一致するかを確認してください.

```
source("http://aoki2.si.gunma-u.ac.jp/R/src/my_t_test.R", encoding="euc-jp")  
my.t.test(50, 69.2, 9.0^2, 150, 60.2, 9.0^2, var.equal=TRUE)
```



2. この患者データから,遺伝子変異ありとなしの群の間に脳萎縮度 (atrophy) の比較を 1. と同じ方法で検定してもよいか?どの検定方法を使えば 1. と同じ検定方法を使えるかどうかを判断できるを説明せよ.実際にこの検定方法を行ってください.

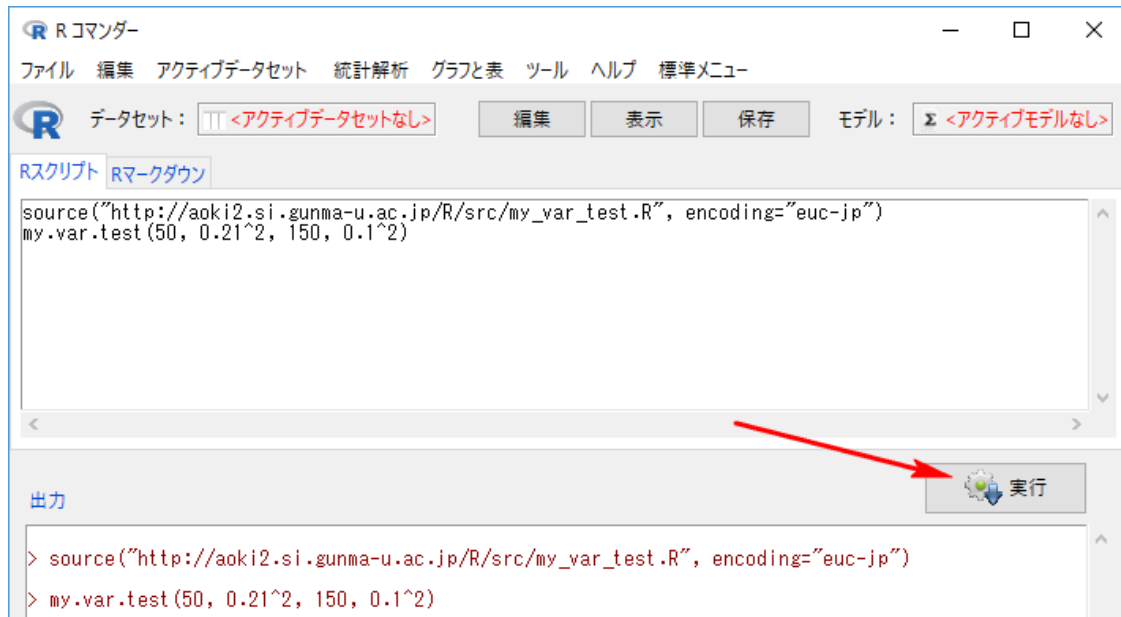
なお,EZR で F 値,両群の分散,両群それぞれの自由度 (df) を使って P 値を計算する時に,以下のコマンドを利用してください:

```
2*pf(F value, df in group 1, df in group 2, lower=FALSE)
```

1.2 答え

以下のコードをRスクリプトに入力して,実行をクリックしてください.自分の検定結果とは一致するかを確認してください.

```
source(" http://aoki2.si.gunma-u.ac.jp/R/src/my_var_test.R" , encoding=" euc-jp" )  
my.var.test(50, 0.21^2, 150, 0.1^2)
```



3. 2.の結果を踏まえて,帰無仮設「両群の脳萎縮度の平均値が等しい」を検定せよ.なお,両群の分散が等しいという前提が満たされていない時に,自由度(df)の計算式は以下となる:

$$df = \frac{(S_A^2/n_A + S_B^2/n_B)^2}{(S_A^2/n_A)^2/(n_A - 1) + (S_B^2/n_B)^2/(n_B - 1)} \quad (2)$$

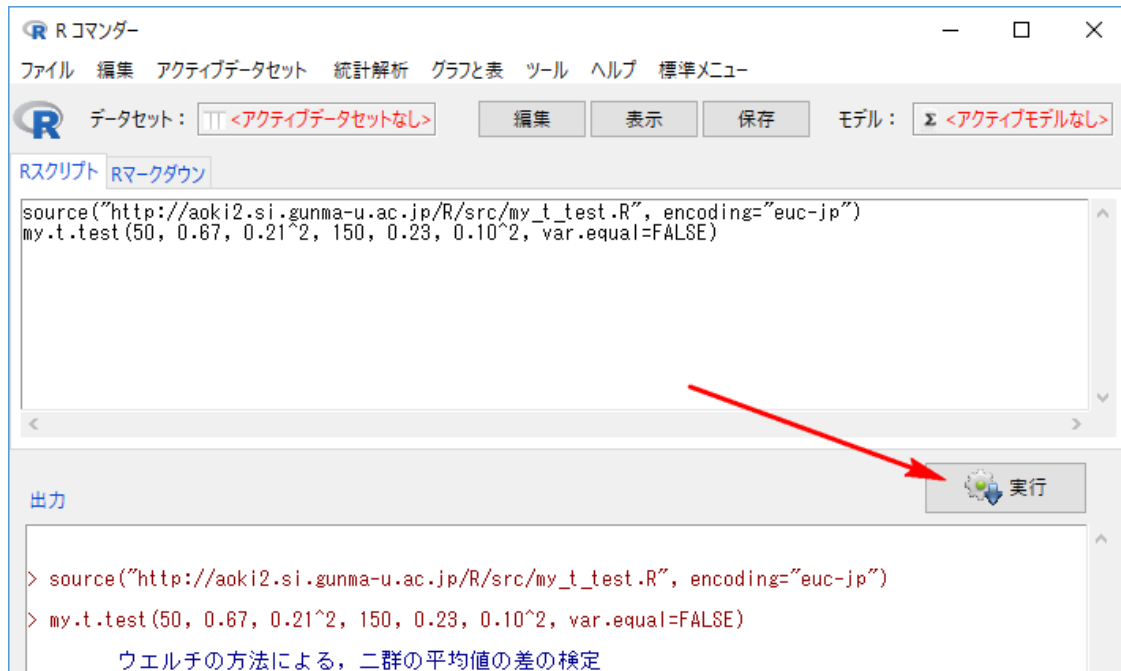
また,EZR で t 値,自由度 (df)を使って P 値を計算する時,以下のコマンドを利用してください:

```
2*pt(t value, df, lower=FALSE)
```

1.3 答え

以下のコードをRスクリプトに入力して,実行をクリックしてください.自分の検定結果とは一致するかを確認してください.

```
source("http://aoki2.si.gunma-u.ac.jp/R/src/my_t_test.R", encoding="euc-jp")  
my.t.test(50, 0.67, 0.21^2, 150, 0.23, 0.10^2, var.equal=FALSE)
```



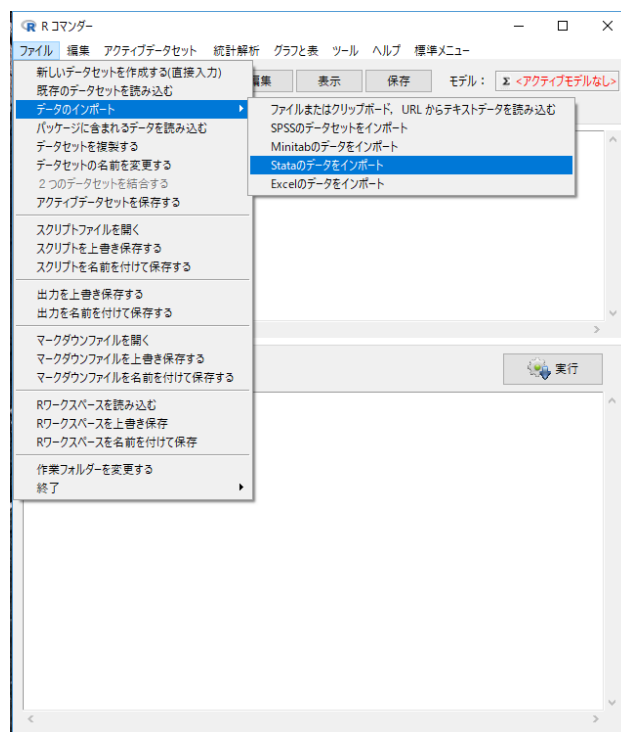
2 問題2:線形回帰モデル

190名の乳幼児の性別(1 = 男,2 = 女),年齢(月, months),体重(kg)のデータを収集した.このデータを用いて,以下の問題を解答したい:

- ・ 子供の年齢が一ヶ月の増加によって,体重はどれぐらい増えているか?
- ・ 男の子は女の子と比べて,平均的に体重はどれぐらい大きい/小さい?

2.1 データのインポート

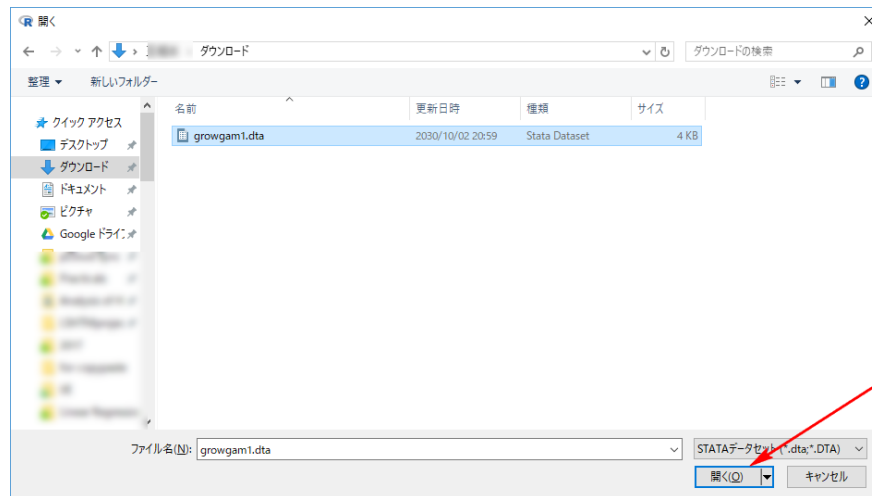
2.1.1 ステップ 1



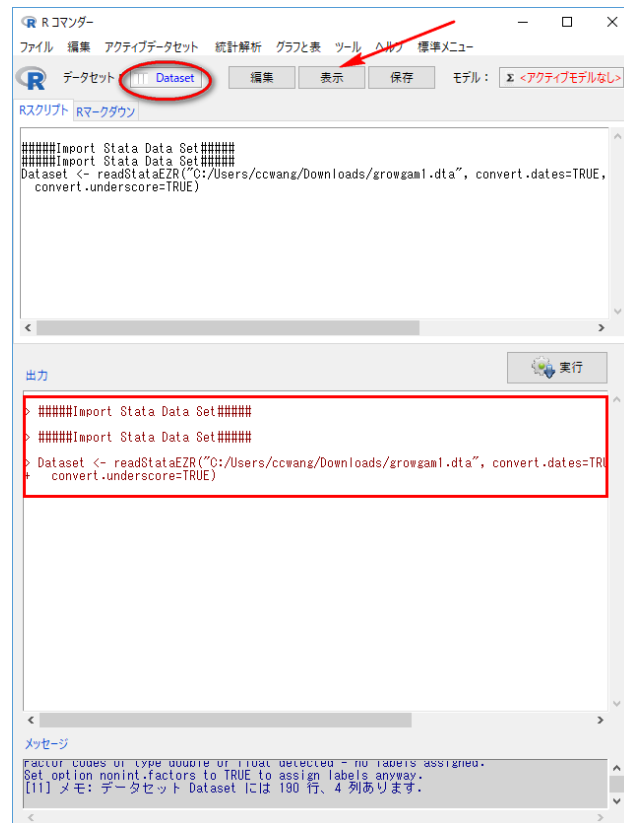
2.1.2 ステップ2



2.1.3 ステップ3



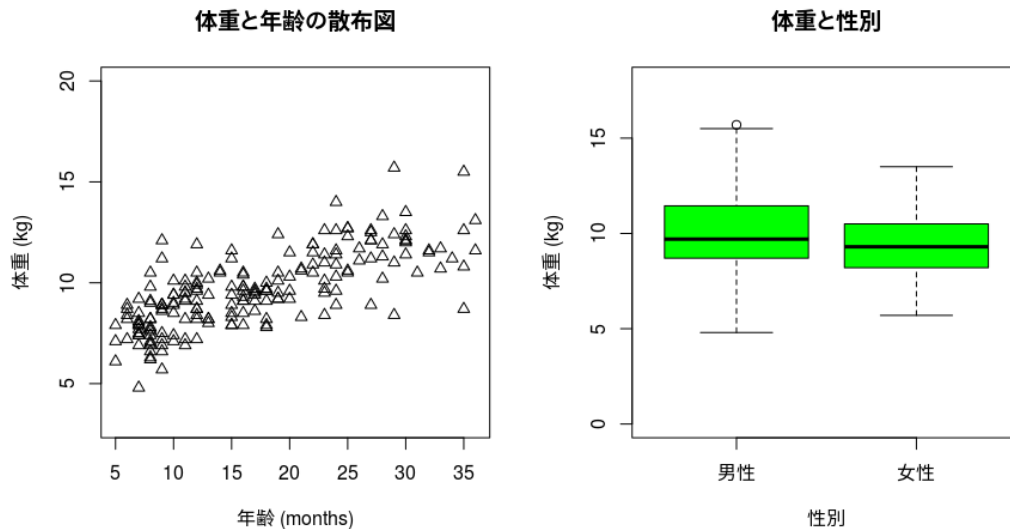
2.1.4 ステップ4



2.1.5 ステップ5

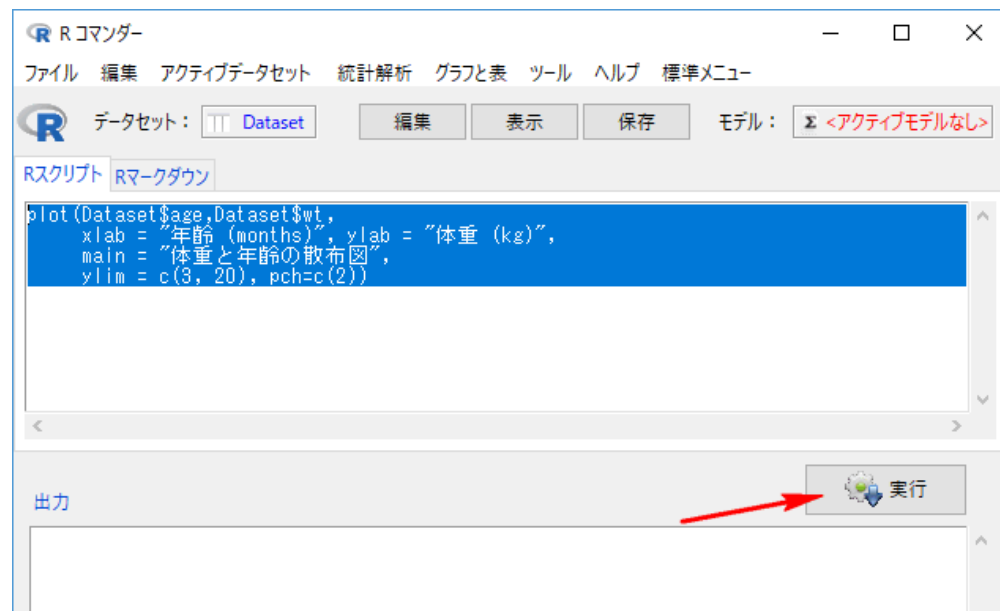
	sex	age	wt	len
1	2	23	8.4	73.2
2	2	22	10.9	84.4
3	2	6	7.2	68.7
4	1	24	10.3	83.7
5	1	14	10.5	79.2
6	2	18	9.6	75.8
7	2	30	11.4	84.4
8	1	24	11.4	84.8
9	1	17	9.4	74.3
10	2	27	12.5	82.6
11	2	18	9.1	74.1
12	1	30	12.0	86.4
13	1	18	7.8	71.9
14	2	15	8.5	77.6
15	2	13	8.0	72.2
16	1	11	9.1	72.4
17	2	8	10.5	71.2
18	2	9	7.2	67.4
19	1	8	6.9	62.7
20	1	16	9.6	79.4
21	2	25	10.5	81.5
22	1	18	9.7	80.7
23	1	29	8.4	81.2
24	2	10	9.4	72.7
25	1	8	7.0	68.3
26	1	9	6.9	68.8
27	1	6	8.7	68.6
28	1	25	12.3	85.4
29	2	16	9.3	78.5
30	1	29	15.7	95.5

2.2 体重と年齢の散布図,性別により体重の箱ひげ図



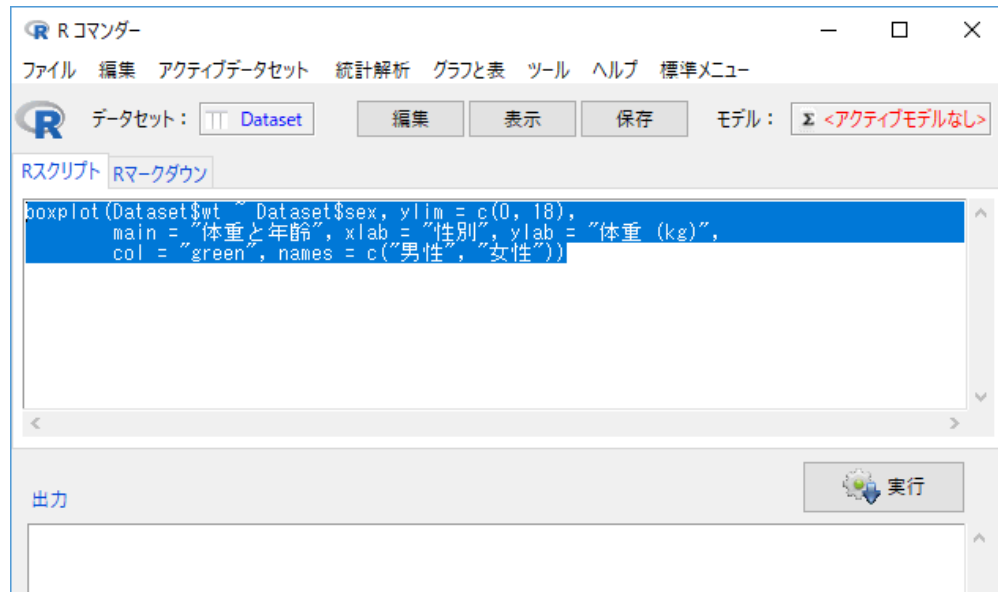
上記左のグラフを描くため,以下のコードをRスクリプトに入力して,実行をクリックしてください。

```
plot(Dataset$age, Dataset$wt,
     xlab = "年齢 (months)", ylab = "体重 (kg)",
     main = "体重と年齢の散布図",
     ylim = c(3, 20), pch=c(2))
```



性別により体重の箱ひげ図を描くため,以下のコードをRスクリプトに入力して,実行をクリックしてください.

```
boxplot(Dataset$wt ~ Dataset$sex, ylim = c(0, 18),
        main = "体重と年齢", xlab = "性別", ylab = "体重 (kg)",
        col = "green", names = c("男性", "女性"))
```



2.3 年齢,体重それぞれの平均値,分散を求めよ;また,年齢と体重の相関係数を算出せよ.なお,EZRで計量データの平均値を計算するには,コマンド `mean(変数名)` を使う;共分散を計算したい時に,コマンド `cor(変数1, 変数2)` を利用する.

以下のコードをRスクリプトに入力して,実行をクリックしてください.(結果を下の余白に記入すること)

```
# 年齢の平均値
mean(Dataset$age)
# 年齢の分散
var(Dataset$age)
# 体重の平均値
mean(Dataset$wt)
# 体重の分散
var(Dataset$wt)
# 体重と年齢の共分散 covariance
cov(Dataset$wt, Dataset$age)
```

2.4 年齢を説明変数, 体重を目的変数とする場合, 年齢の傾き(回帰係数), と切片を求めよ. なお, 分散と共分散の定義を以下とする, \bar{X} は X の平均値を示す:

- 分散 variance:

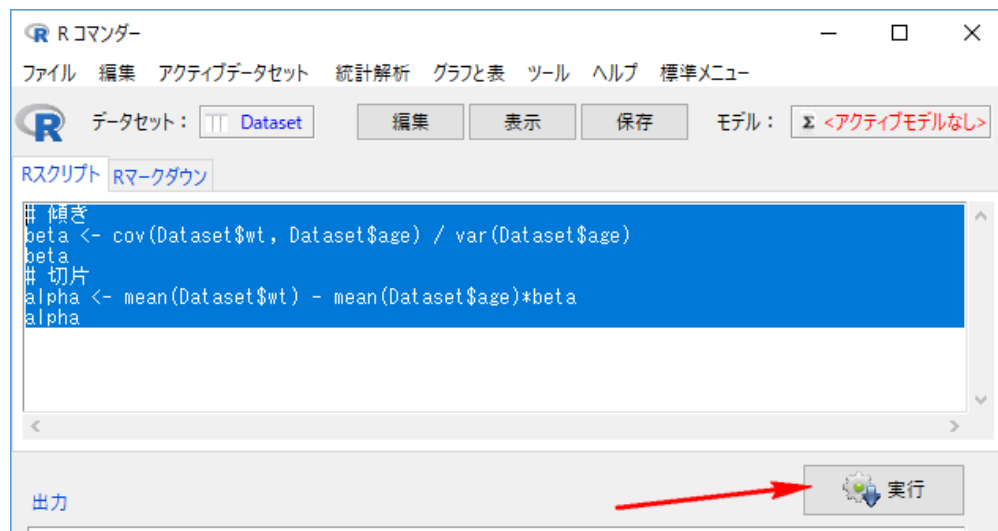
$$\begin{aligned}\text{Var}(X) &= \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n - 1} \\ &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}\end{aligned}$$

- 共分散 covariance:

$$\begin{aligned}\text{Cov}(X, Y) &= \frac{(X_1 - \bar{X})(Y_1 - \bar{Y}) + (X_2 - \bar{X})(Y_2 - \bar{Y}) + \cdots + (X_n - \bar{X})(Y_n - \bar{Y})}{n - 1} \\ &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}\end{aligned}$$

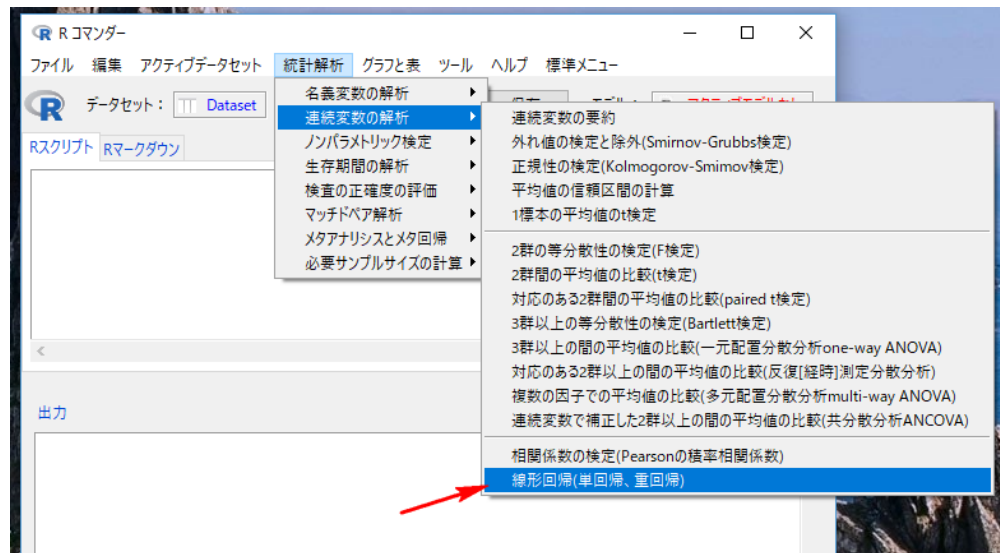
以下のコードをRスクリプトに入力して, 実行をクリックしてください. (結果を下の余白に記入すること)

```
# 傾き (slope)
beta <- cov(Dataset$wt, Dataset$age) / var(Dataset$age)
beta
# 切片 (intercept)
alpha <- mean(Dataset$wt) - mean(Dataset$age)*beta
alpha
```

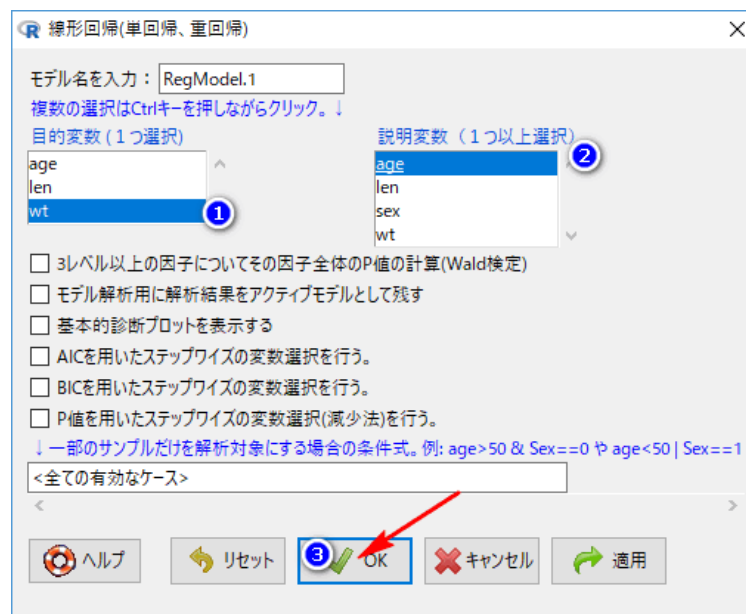


2.5 実際にEZRで線形モデルを作ってみよう:

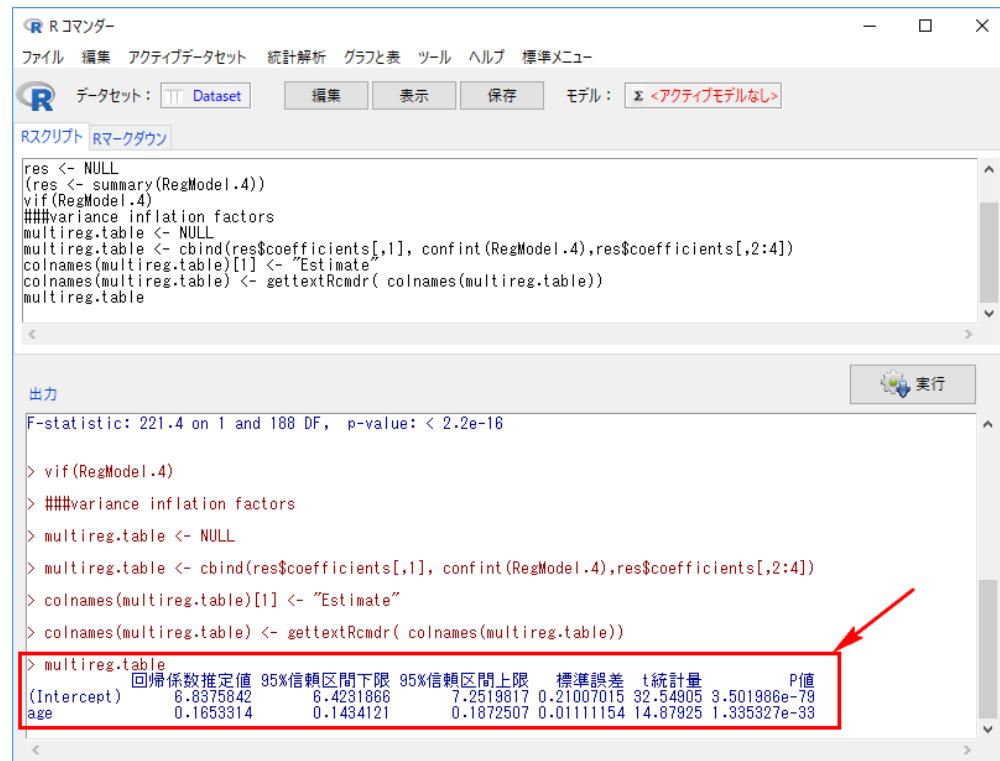
2.5.1 ステップ1



2.5.2 ステップ2



2.5.3 ステップ3



```

R コマンドー
ファイル 編集 アクティブデータセット 統計解析 グラフと表 ツール ヘルプ 標準メニュー

データセット: TT Dataset 編集 表示 保存 モデル: <アクティブモデルなし>

Rスクリプト Rマークダウン

res <- NULL
(res <- summary(RegModel.4))
vif(RegModel.4)
###variance inflation factors
multireg.table <- NULL
multireg.table <- cbind(res$coefficients[,1], confint(RegModel.4), res$coefficients[,2:4])
colnames(multireg.table)[1] <- "Estimate"
colnames(multireg.table) <- gettextRcmdr( colnames(multireg.table))
multireg.table

出力
F-statistic: 221.4 on 1 and 188 DF, p-value: < 2.2e-16

> vif(RegModel.4)
> ###variance inflation factors
> multireg.table <- NULL
> multireg.table <- cbind(res$coefficients[,1], confint(RegModel.4), res$coefficients[,2:4])
> colnames(multireg.table)[1] <- "Estimate"
> colnames(multireg.table) <- gettextRcmdr( colnames(multireg.table))
> multireg.table

```

	回帰係数推定値	95%信頼区間下限	95%信頼区間上限	標準誤差	t統計量	P値
(Intercept)	6.8375842	6.4231866	7.2519817	0.21007015	32.54905	3.501986e-79
age	0.1653314	0.1434121	0.1872507	0.01111154	14.87925	1.335327e-33

自分の計算結果とは一致するかを確認してください。

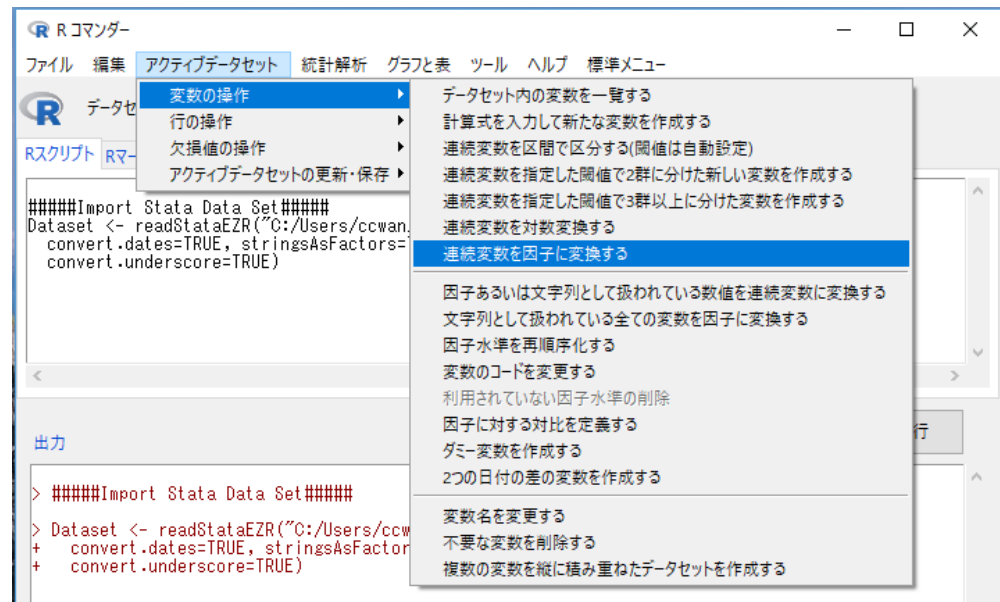
2.6 今まで計算した傾きと切片の数字を用いて、年齢と体重の関係を線形と考える場合の計算式を記入せよ。傾きと切片の計算結果の意味をそれぞれ記述せよ。

2.6.1 答え

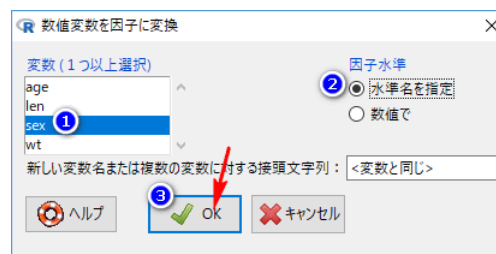
2.7 性別を説明変数に入れたモデルを作る

2.7.1 性別変数を因子 (factor) に変換する

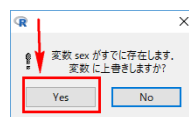
2.7.1.1 ステップ1



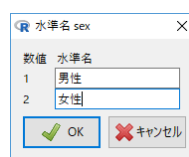
2.7.1.2 ステップ2



2.7.1.3 ステップ3

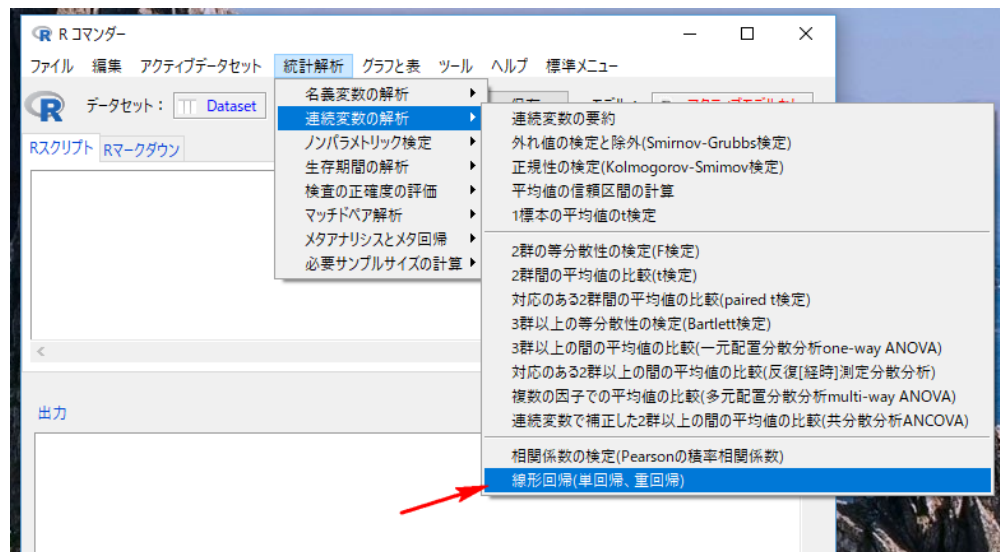


2.7.1.4 ステップ4-水準名に男性,女性を入力する

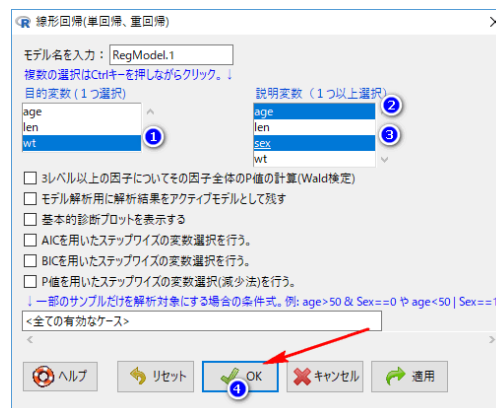


2.7.2 重回帰線形モデルを作る

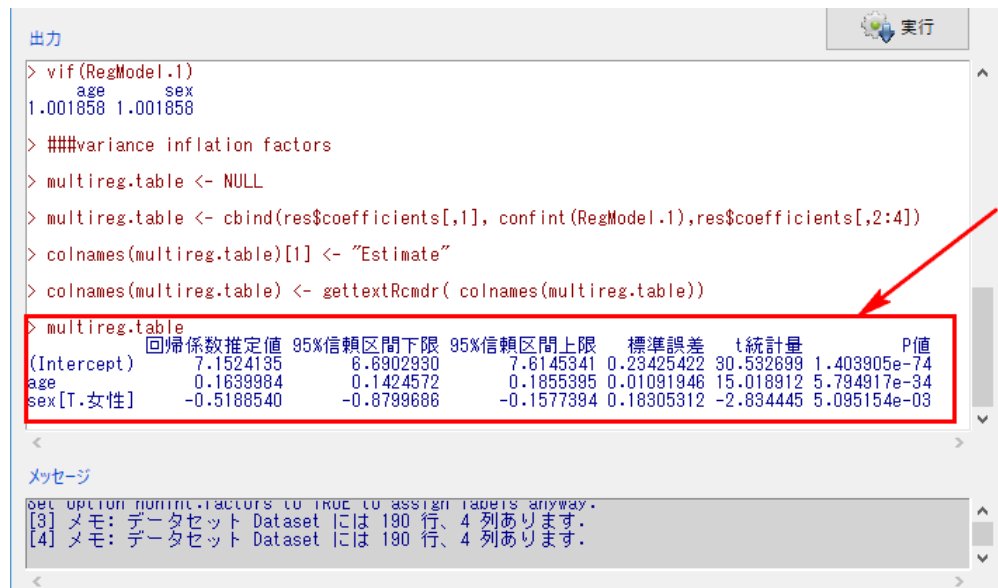
2.7.2.1 ステップ1



2.7.2.2 ステップ2—複数の説明変数を選択する時に control キーを押しながらマウスで変数名をクリックする



2.7.3 重回帰線形モデルの結果を確認する



```

出力
> vif(RegModel.1)
      age      sex 
1.001858 1.001858 

> ###variance inflation factors

> multireg.table <- NULL

> multireg.table <- cbind(res$coefficients[,1], confint(RegModel.1), res$coefficients[,2:4])

> colnames(multireg.table)[1] <- "Estimate"

> colnames(multireg.table) <- gettextRcmdr( colnames(multireg.table))

> multireg.table
      回帰係数推定値 95%信頼区間下限 95%信頼区間上限 標準誤差  t統計量  P値
(Intercept)      7.1524135         6.8902930       7.6145341  0.23425422 30.532699 1.403905e-74
age              0.1639984         0.1424572       0.1855395  0.01091946 15.018912 5.794917e-34
sex[T.女性]      -0.5188540        -0.8799686      -0.1577394  0.18305312 -2.834445 5.095154e-03

メッセージ
set OPTION nomint.factors to TRUE to assign labels anyway.
[3] メモ: データセット Dataset には 190 行、4 列あります。
[4] メモ: データセット Dataset には 190 行、4 列あります。

```

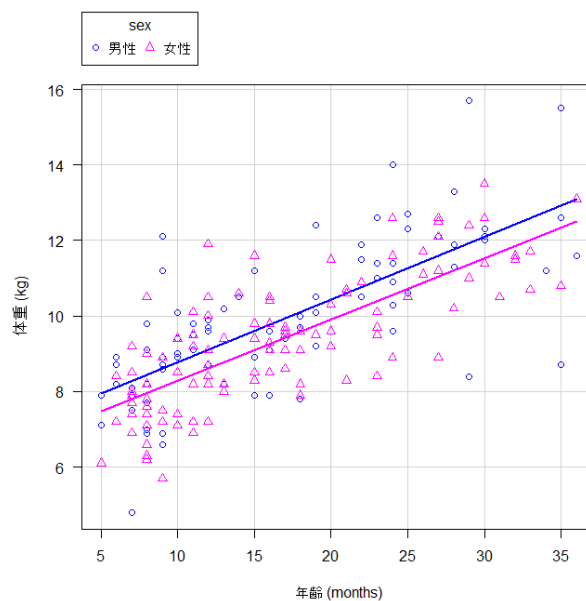
2.8 重回帰線形モデルの計算結果を用いて、体重の平均値を年齢と性別の線形モデルで表示せよ。各回帰係数の意味を説明せよ。

2.9 答え

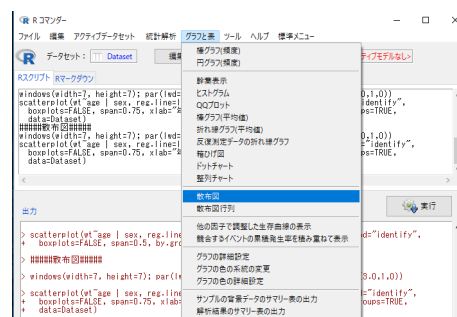
2.10 上記の重回帰線形モデルを用いて、年齢が34ヶ月の女の子の体重の予測値を計算せよ。

2.11 答え

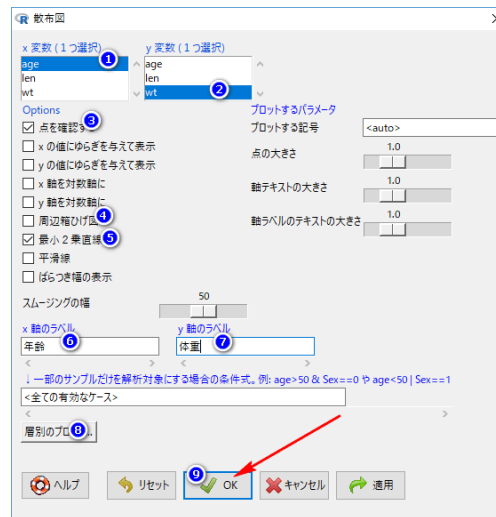
2.12 男女別の年齢と体重の散布図を描く



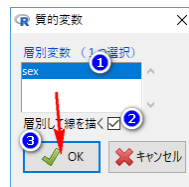
2.12.1 ステップ1



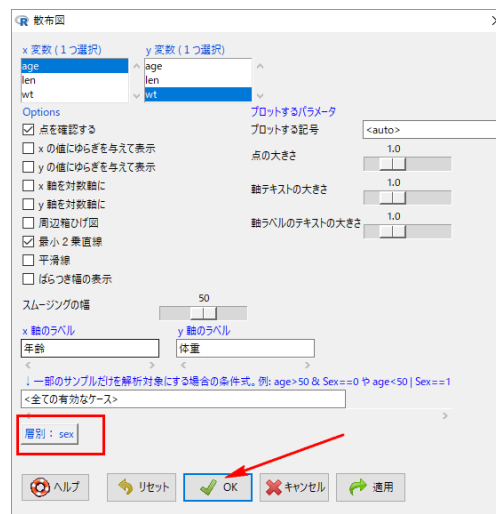
2.12.1.1 ステップ2



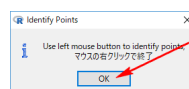
2.12.1.2 ステップ3



2.12.1.3 ステップ4



2.12.1.4 ステップ5



3 参考図書:

- 1.「Rによる保健医療データ解析演習」,中澤 港, (<http://minato.sip21c.org/msb/medstatbookx.pdf>)
- 2.「みんなの医療統計 12日間で基礎理論とEZRを完全マスター!」,新谷 歩.