

# 統計分析法 第8週レポート

202212022 田島瑞起

2023/12/12

## 1 (7) 解答

交互作用も考慮した場合の重回帰分析モデルを下に表す。  $grmax = -27.12 + 0.1797 * age + 0.3545 * ht - 2.36 * wt - 0.001728 * (age * ht) + 0.01879 * (age * wt) + 0.01614 * (ht * wt) - 0.0001217 * (age * ht * wt)$  自由度調整済み決定係数に関しては、Multiple R-squared:0.6138, Adjusted R-squared:0.6108 であり、回帰残差は 6.052 となる。

```
1 Call:
2 lm(formula = grmax ~ age * ht * wt, data = data)
3
4 Residuals:
5     Min       1Q   Median       3Q      Max
6  -28.920  -3.795  -0.262   3.627  22.681
7
8 Coefficients:
9             Estimate Std. Error t value Pr(>|t|)
10 (Intercept) -2.712e+01  1.182e+02  -0.230   0.819
11 age          1.797e-01  1.882e+00   0.095   0.924
12 ht           3.545e-01  7.549e-01   0.470   0.639
13 wt          -2.360e+00  1.999e+00  -1.181   0.238
14 age:ht       -1.728e-03  1.208e-02  -0.143   0.886
15 age:wt       1.879e-02  3.244e-02   0.579   0.563
16 ht:wt        1.614e-02  1.258e-02   1.282   0.200
17 age:ht:wt    -1.217e-04  2.051e-04  -0.593   0.553
18
19 Residual standard error: 6.052 on 901 degrees of freedom
20 Multiple R-squared:  0.6138,    Adjusted R-squared:  0.6108
21 F-statistic: 204.6 on 7 and 901 DF,  p-value: < 2.2e-16
```

(5) の結果と比較するために、(5) の summary を下に表す。

```
1 Call:
2 lm(formula = grmax ~ age + ht + wt, data = data)
3
4 Residuals:
5     Min       1Q   Median       3Q      Max
6  -28.9156  -3.9633  -0.2197   3.7000  22.3128
7
8 Coefficients:
9             Estimate Std. Error t value Pr(>|t|)
10 (Intercept) -88.46022    4.59757  -19.241  < 2e-16 ***
11 age         -0.10569    0.02027   -5.213  2.30e-07 ***
12 ht           0.75014    0.03132   23.950  < 2e-16 ***
13 wt           0.16476    0.02665    6.182  9.54e-10 ***
14 ---
15 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
16
17 Residual standard error: 6.181 on 905 degrees of freedom
18 Multiple R-squared:  0.5955,    Adjusted R-squared:  0.5941
19 F-statistic: 444.1 on 3 and 905 DF,  p-value: < 2.2e-16
```

総合的に、Adjusted R-squared が高く、残差も小さい交互作用を考慮したモデルの方が、(5) のモデルよりも実態に即していると言える可能性がある。しかし p 値を見ると (7) のモデルはどれも有意水準よりもはるかに高く、効果がないという可能性

もある点に注意する必要がある。

## 2 (8) 解答

step 関数を使用し、(7) で求めた AIC より小さい値を保有するモデルを算出すると、下記のモデルが考えられた。

```
1 Call:
2 lm(formula = grmax ~ age + ht + wt + age:ht + ht:wt, data = data)
3
4 Residuals:
5      Min       1Q   Median       3Q      Max
6 -29.1085  -3.7875  -0.2145   3.6634  22.6312
7
8 Coefficients:
9             Estimate Std. Error t value Pr(>|t|)
10 (Intercept) -93.703024   38.073569  -2.461 0.014037 *
11 age          1.286273    0.380180   3.383 0.000747 ***
12 ht           0.786496    0.243723   3.227 0.001296 **
13 wt          -1.238333    0.383907  -3.226 0.001302 **
14 age:ht       -0.008927    0.002431  -3.672 0.000255 ***
15 ht:wt        0.008889    0.002436   3.649 0.000279 ***
16 ---
17 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
18
19 Residual standard error: 6.047 on 903 degrees of freedom
20 Multiple R-squared:  0.6136,    Adjusted R-squared:  0.6115
21 F-statistic: 286.8 on 5 and 903 DF,  p-value: < 2.2e-16
```

このモデルを数式で表すと下記の通りとなる。 $grmax = -93.703024 + 1.28627 * age + 0.78649 * ht - 1.23833 * wt - 0.008927 * age * ht + 0.008889 * ht * wt$  また、 $AiC = 5859.224$  である。

## 3 (9) 解答

```
1 Call:
2 lm(formula = grmax ~ wt, data = Data)
3
4 Residuals:
5      Min       1Q   Median       3Q      Max
6 -34.854  -5.470  -0.692   4.808  27.145
7
8 Coefficients:
9             Estimate Std. Error t value Pr(>|t|)
10 (Intercept) -0.91185    1.67601  -0.544  0.587
11 wt           0.55541    0.02853  19.469 <2e-16 ***
12 ---
13 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
14
15 Residual standard error: 8.152 on 907 degrees of freedom
16 Multiple R-squared:  0.2947,    Adjusted R-squared:  0.294
17 F-statistic: 379 on 1 and 907 DF,  p-value: < 2.2e-16
```

```
1 Call:
2 lm(formula = grmax ~ ht, data = Data)
3
4 Residuals:
5      Min       1Q   Median       3Q      Max
6 -25.9019  -4.0089  -0.2739   3.9494  23.2542
7
8 Coefficients:
9             Estimate Std. Error t value Pr(>|t|)
10 (Intercept) -106.9868    4.0168  -26.64 <2e-16 ***
11 ht           0.8893     0.0258   34.47 <2e-16 ***
12 ---
```

```

13 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
14
15 Residual standard error: 6.386 on 907 degrees of freedom
16 Multiple R-squared:  0.5671,    Adjusted R-squared:  0.5667
17 F-statistic: 1188 on 1 and 907 DF,  p-value: < 2.2e-16

```

```

1 Call:
2 lm(formula = grmax ~ age, data = Data)
3
4 Residuals:
5      Min       1Q   Median       3Q      Max
6 -19.972  -7.061  -2.351   6.678  30.290
7
8 Coefficients:
9             Estimate Std. Error t value Pr(>|t|)
10 (Intercept)  46.30143     1.85535   24.956 < 2e-16 ***
11 age         -0.24630     0.03001   -8.206 7.77e-16 ***
12 ---
13 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
14
15 Residual standard error: 9.366 on 907 degrees of freedom
16 Multiple R-squared:  0.06912,    Adjusted R-squared:  0.06809
17 F-statistic: 67.35 on 1 and 907 DF,  p-value: 7.769e-1

```

単回帰の結果をそれぞれ読み取ると、grmax は wt に強い正の相関を持ち、ht に弱い正の相関、age に弱い負の相関を持つことが読み取れる。また先ほどの交互作用項を考慮していない重回帰モデルでは、ht に強い相関があることが分かる。また交互作用項を考慮した場合の重回帰モデルで、決定係数及、残差標準偏差、AIC 全てにおいて最適にモデルを示していると考えられる (8) で作成したモデルを見ると、age, ht 項に正の相関が存在し、wt に負の相関が生じた。それぞれの分析法によって変数間の関係性が異なるが、最もモデルを反映している改良済みの交互作用項を考慮した重回帰分析の結果を重視すると、交互作用項による影響は wt, ht が共に高い場合 grmax が増加すると推測が出来る。また age, ht に正の相関があり、wt に負の相関があることが読み取れる。最終的な判断としては唯体重が重いだけでは握力に負の影響を及ぼすが、身長と年齢は単独で grmax に正の影響を及ぼし、身長に付随して体重も重い状態であれば握力に正の影響を及ぼすと考えられる。

## 4 ソースコード

```

1 Data <- read.table("week9-data.txt", header = TRUE)
2 # データの整形
3 age <- Data$age
4 ht <- Data$ht
5 wt <- Data$wt
6 grmax <- Data$grmax
7
8 # 単回帰分析
9 model_age <- lm(grmax ~ age, data = Data)
10 model_ht <- lm(grmax ~ ht, data = Data)
11 model_wt <- lm(grmax ~ wt, data = Data)
12
13 # 結果の表示
14 summary(model_age)
15 summary(model_ht)
16 summary(model_wt)
17
18 # データフレームの用意
19 data <- data.frame(age = age,
20                   ht = ht,
21                   wt = wt,
22                   grmax = grmax)
23
24 # 重回帰モデルの構築
25 model <- lm(grmax ~ age + ht + wt, data = data)
26
27 # モデルのサマリーを表示

```

```

28 summary(model)
29
30 data <- data.frame(age = age,
31                   ht = ht,
32                   wt = wt)
33
34 # 相関行列の計算
35 cor_matrix <- cor(data)
36
37 # 相関行列の表示
38 print(cor_matrix)
39
40 # 重回帰モデルに交互作用項を含める
41 model_interaction <- lm(grmax ~ age * ht * wt, data = data)
42
43 # モデルのサマリーを表示
44 summary(model_interaction)
45
46 aic_value <- AIC(model_interaction)
47
48 # step関数を使用して変数の選択を行う
49 final_model <- step(model_interaction , direction = "both", trace = 0)
50
51 # ファイナルモデルの表示
52 summary(final_model)
53
54 aic_value <- AIC(final_model)

```