

Лабораторна робота № 5

Завдання: Для множинної лінійної регресійної моделі представити прогноз та довірчі інтервали.

(A) Побудувати підсумкову функцію ANOVA для функції за 5-ма параметрами та вказати:

```
> # Залежна змінна rating
> Y <- data$rating
> x1 <- data$calories
> x2 <- data$sodium
> x3 <- data$protein
> x4 <- data$potass
> x5 <- data$cup
> mod <- lm(Y ~ x1+x2+x3+x4+x5)
> simpleAnova(mod)
Analysis of Variance Table

Response: Y
      Df Sum Sq Mean Sq F value    Pr(>F)
Predictors  5 11214.1  2242.82  42.097 < 2.2e-16 ***
Residuals  71  3782.7    53.28
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a. SSR для p змінних;

Для $p = 5$, $SSR = 2242.82$

b. SSE для $n - p - 1$ степенів вільності;

Для $77 - 5 - 1 = 71$: 53.28

c. F- статистику;

F- статистика = 42.097

(В) Побудувати функцією `anova(*)` за вхідними 5-ма параметрами;

```
> anova(mod)
Analysis of Variance Table

Response: Y
      Df Sum Sq Mean Sq  E value    Pr(>F)
x1      1 7127.1   7127.1  133.7718 < 2.2e-16 ***
x2      1  620.7    620.7   11.6508  0.001064 **
x3      1 3339.8   3339.8   62.6860 2.379e-11 ***
x4      1  111.5    111.5    2.0929  0.152384
x5      1   15.0     15.0    0.2815  0.597351
Residuals 71 3782.7    53.3
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a. Вказати $SSR_1, SSR_2, SSR_3, SSR_4, SSR_5$;

$$SSR_1 = 7127.1$$

$$SSR_2 = 620.7$$

$$SSR_3 = 3339.8$$

$$SSR_4 = 111.5$$

$$SSR_5 = 15.0$$

b. Вказати F-value для 5-ти параметрів;

$$F_1 = 133.77$$

$$F_2 = 11.65$$

$$F_3 = 62.68$$

$$F_4 = 2.09$$

$$F_5 = 0.28$$

c. Вказати ті які не проходять за F-статистикою;

x_4, x_5 – незначущі, вони не проходять за F-статистикою.

(C) Зробити покращення за допомогою AIC/BIC моделі за всіма параметрами які є в dataset.

a. Побудувати `MASS::stepAIC(*, direction = "backward", k = log(n))`. Описати фінальну модель;

```
Step: AIC=-552.62
rating ~ calories + protein + fat + sodium + fiber + carbo +
sugars + potass + vitamins + weight

Df Sum of Sq    RSS    AIC
<none>                0.03 -552.62
- weight      1      0.00    0.03 -552.24
- potass      1    50.16   50.19   10.49
- fat         1    66.71   66.74   32.43
- vitamins    1    79.35   79.38   45.78
- sugars      1   124.01  124.04   80.15
- calories    1   139.22  139.25   89.06
- carbo       1   335.78  335.81  156.84
- protein     1   380.29  380.32  166.42
- fiber       1   605.15  605.18  202.19
- sodium      1  1177.55 1177.59  253.45

Call:
lm(formula = rating ~ calories + protein + fat + sodium + fiber +
carbo + sugars + potass + vitamins + weight, data = data)

Coefficients:
(Intercept)      calories      protein         fat      sodium      fiber      carbo      sugars      potass      vitamins
54.88246      -0.22565      3.28465     -1.66201     -0.05450      3.45394      1.11008     -0.70895     -0.03416     -0.05127
weight
-0.07519
```

Було видалено 2 параметри: спочатку cups, а потім shelf. Так, як ці параметри є неважливими. Step AIC = -552.62

b. `MASS::stepAIC(modZero, direction = "forward", scope = list(lower = modZero, upper = modAll), k = log(n))` Описати фінальну модель;

```
Step: AIC=-552.62
rating ~ sugars + fiber + sodium + fat + protein + carbo + calories +
vitamins + potass + weight

Df Sum of Sq    RSS    AIC
<none>                0.031630 -552.62
+ shelf  1  0.0016787  0.029951 -552.48
+ cups   1  0.0012126  0.030417 -551.29

Call:
lm(formula = rating ~ sugars + fiber + sodium + fat + protein +
carbo + calories + vitamins + potass + weight, data = data)

Coefficients:
(Intercept)      sugars      fiber      sodium      fat      protein      carbo      calories      vitamins      potass
54.88246      -0.70895      3.45394     -0.05450     -1.66201      3.28465      1.11008     -0.22565     -0.05127     -0.03416
weight
-0.07519
```

Алгоритм йшов від моделі без факторів поступово додаючи нові. Бачимо, що додалися усі, окрім shelf та cups.

c. `MASS::stepAIC(modAll, direction = "both", trace = 0, scope = list(lower = modZero, upper = modAll), k = log(n))`

```
> ### c.
> MASS::stepAIC(modAll, direction = "both", trace = 0, scope = list(lower = modZero, upper = modAll), k = log(nrow(data)))

Call:
lm(formula = rating ~ calories + protein + fat + sodium + fiber +
carbo + sugars + potass + vitamins + weight, data = data)

Coefficients:
(Intercept)      calories      protein         fat      sodium      fiber      carbo      sugars      potass      vitamins
54.88246      -0.22565      3.28465     -1.66201     -0.05450      3.45394      1.11008     -0.70895     -0.03416     -0.05127
weight
-0.07519
```

d. Побудувати критерій AIC MASS::stepAIC(*, trace = 0, k = 2). Описати фінальну модель;

```
> AIC <- MASS::stepAIC(modAll, trace = 0, k = 2)
> AIC

call:
lm(formula = rating ~ calories + protein + fat + sodium + fiber +
    carbo + sugars + potass + vitamins + shelf + weight, data = data)

Coefficients:
(Intercept)      calories      protein         fat      sodium      fiber      carbo      sugars      potass      vitamins
  54.869987    -0.225613     3.285137    -1.663954    -0.054488     3.453535     1.109762    -0.709158    -0.034180    -0.051369
    shelf      weight
   0.006847    -0.071252
```

Критерій AIC має усі параметри, окрім cups.

e. Побудувати критерій BIC MASS::stepAIC(*, trace = 0, k = log(n))). Описати фінальну модель;

```
> BIC <- MASS::stepAIC(modAll, trace = 0, k = log(nrow(data)))
> BIC

call:
lm(formula = rating ~ calories + protein + fat + sodium + fiber +
    carbo + sugars + potass + vitamins + weight, data = data)

Coefficients:
(Intercept)      calories      protein         fat      sodium      fiber      carbo      sugars      potass      vitamins
  54.88246    -0.22565     3.28465    -1.66201    -0.05450     3.45394     1.11008    -0.70895    -0.03416    -0.05127
    weight
  -0.07519
```

Критерій BIC має усі параметри, окрім cups і shelf.