

Лабораторна робота № 6

Завдання:

(А) Побудуйте:

```
mod1 <- lm(Y ~ ., data)
mod2 <- lm(Y ~ . - x, data)
mod3 <- lm(Y ~ x1 + x2, data)
> mod1 <- lm(rating ~ ., data = data)
> mod2 <- lm(rating ~ . - cups, data = data)
> mod3 <- lm(rating ~ calories + potass, data = data)
```

(В) Аналіз 3-х уточнюючих моделей з попереднього пункту за відповідними значеннями R^2 та R^2_{Adj} та вказати яка краща;

```
> summary(mod1)

Call:
lm(formula = rating ~ ., data = data)

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02141 on 64 degrees of freedom
Multiple R-squared:  1,      Adjusted R-squared:  1
F-statistic: 2.726e+06 on 12 and 64 DF,  p-value: < 2.2e-16

> summary(mod2)

Call:
lm(formula = rating ~ . - cups, data = data)

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02147 on 65 degrees of freedom
Multiple R-squared:  1,      Adjusted R-squared:  1
F-statistic: 2.959e+06 on 11 and 65 DF,  p-value: < 2.2e-16

> summary(mod3)

Call:
lm(formula = rating ~ calories + potass, data = data)

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.141 on 74 degrees of freedom
Multiple R-squared:  0.5877,   Adjusted R-squared:  0.5765
F-statistic: 52.74 on 2 and 74 DF,  p-value: 5.806e-15
```

У моделі 1 і моделі 2 $R^2 = R^2_{Adj} = 1$, що дуже добре. І вони краще ніж модель 3 у якій $R^2 = 0.5877$ і $R^2_{Adj} = 0.5765$.

(С) Побудувати додаткову модуль без перетину та зробити висновки для R^2 та R^2_{Adj}

```
m1 <- lm(rating ~ calories, data = data)
m0 <- lm(rating ~ 0 + calories, data = data)
```

M1 – з перетином

M0 – без перетину

```
> summary(m1)
```

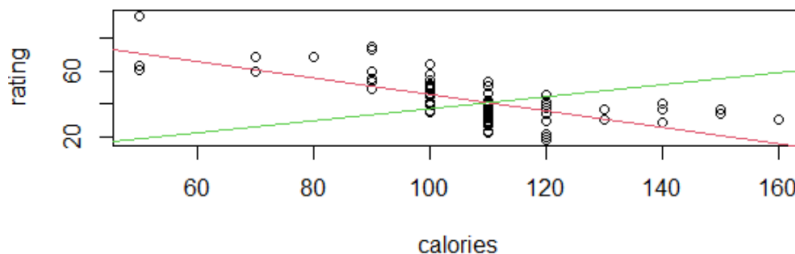
```
Call:
lm(formula = rating ~ calories, data = data)
```

Residual standard error: 10.24 on 75 degrees of freedom
Multiple R-squared: 0.4752, Adjusted R-squared: 0.4682
F-statistic: 67.92 on 1 and 75 DF, p-value: 4.14e-12

```
> summary(m0)
```

```
Call:
lm(formula = rating ~ 0 + calories, data = data)
```

Residual standard error: 19.97 on 76 degrees of freedom
Multiple R-squared: 0.8047, Adjusted R-squared: 0.8021
F-statistic: 313.1 on 1 and 76 DF, p-value: < 2.2e-16



Хоча у моделі без перетину набагато краще $R^2 = 0.8047$, ніж у моделі з перетином $R^2 = 0.4752$ та $R^2_{Adj} = 0.8021$ краще $R^2_{Adj} = 0.4682$. Але модель з перетином набагато краща на графіку.

(D) Побудувати центровану модель та зробити висновки для R^2 та R^2_{Adj} .

```
dataCen <- data.frame(scale(data, center = TRUE, scale = FALSE))
modCen1 <- lm(rating ~ calories, data = dataCen)
modCen0 <- lm(rating ~ 0 + calories, data = dataCen)
```

```
> summary(modCen1)
```

```
Call:
lm(formula = rating ~ calories, data = dataCen)
```

Residuals:

Min	1Q	Median	3Q	Max
-18.7201	-7.9317	-0.6678	5.9902	23.4161

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.420e-15	1.167e+00	0.000	1
calories	-4.970e-01	6.031e-02	-8.241	4.14e-12 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.24 on 75 degrees of freedom
Multiple R-squared: 0.4752, Adjusted R-squared: 0.4682
F-statistic: 67.92 on 1 and 75 DF, p-value: 4.14e-12

```
> summary(modCen0)
```

```
Call:
lm(formula = rating ~ 0 + calories, data = dataCen)

Residuals:
    Min       1Q   Median       3Q      Max
-18.7201  -7.9317  -0.6678   5.9902  23.4161

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
calories -0.49701    0.05991   -8.296 2.98e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.18 on 76 degrees of freedom
Multiple R-squared:  0.4752,    Adjusted R-squared:  0.4683
F-statistic: 68.83 on 1 and 76 DF,  p-value: 2.98e-12
```

$R^2 = 0.4752$ і $R_{Adj}^2 = 0.4683$ у обох центрованих моделей (з перетином і без). Також такі ж самі R^2 і R_{Adj}^2 має нецентрована модель з перетином.