In class exercise week 9
Topic: linear regression

A researcher examines the influence of several possible explanatory variables to an enzyme which is essential for the survival of cricket eggs (Grilleneier) and which can be inhibited by insecticides. The data set `cricket` contains $n = 156$ observations, each corresponding to a measurement in one egg, of the following variables:

`activity`	Activity of the enzyme in the egg
`treated`	Age of the egg (in days) at the time of the insecticide treatment
`observed`	Age of the egg (in days) at the time of the enzyme activity measurement
`insecticide`	Type of insecticide: carbaryl (0) or propoxur (1)
`dosage`	Dosage of the insecticide: low (0.6 mg), medium (0.8 mg), high (1.0 mg)

The researcher fitted a simple linear regression with `treated` is the only considered variable explaining the activity of investigated eggs; this yielded the following output:

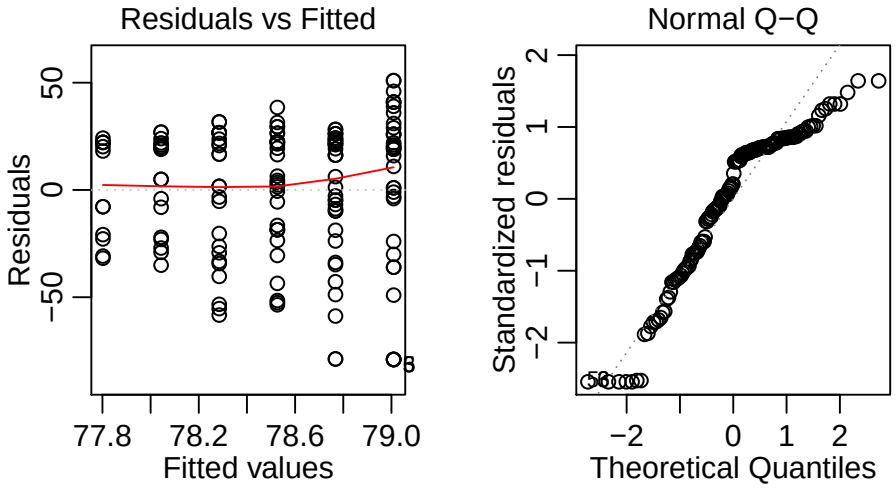
```
Call:
lm(formula = activity ~ treated, data = cricket)

Residuals:
    Min      1Q  Median      3Q     Max
-79.009 -22.233   8.732  22.663  50.991

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  79.7331     8.2626   9.650   <2e-16 ***
treated      -0.2414     1.5993  -0.151     0.88
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 31.3 on 154 degrees of freedom
Multiple R-squared: 0.0001479,        Adjusted R-squared: -0.006345
F-statistic: 0.02278 on 1 and 154 DF,  p-value: 0.8802
```

Suppose that the model assumptions on a linear model are satisfied when answering the following two questions.

i) The fitted linear model is not significantly better than a model which only takes the mean activity of the eggs as prediction on a 5% level.
☐ True ☐ False

ii) The intercept is different of 70 on a significance level of 5%.
☐ True ☐ False

We now analyze the residuals. The following three questions refer to them:



iii) The model assumptions on the expectation values of the errors are clearly violated.
☐ True ☐ False

iv) The distribution of the residuals has a shorter tail than a normal distribution.
☐ True ☐ False

v) A log-transformation could help to stabilize the variance of the residuals. ☐ True ☐ False