Statistics Methods in Finance
Homework 5

0786009 廖家鴻

DUE 2020/12/2 00:00

Outline (HW5 questions)

1. (50%) Find some appropriate nonlinear function with two parameters, and do a nonlinear regression (*Hint. See the introduction pdf file on E3*.)

2. (50%) Do a hypothesis testing on the coefficients, both through linearization and through likelihood approach.

1. Find some nonlinear function with two parameters

Example e **Biochemical Oxygen Demand** To determine the biochemical oxygen demand, stream water samples were enriched with soluble organic matter, with inorganic nutrients and with dissolved oxygen, and subdivided into bottles (Marske, 1967, see Bates and Watts, 1988). Each bottle was inoculated with a mixed culture of microorganisms, sealed and put in a climate chamber with constant temperature. The bottles were periodically opened and their dissolved oxygen concentration was analyzed, from which the biochemical oxygen demand [mg/l] was calculated. The model used to connect the cumulative biochemical oxygen demand Y with the incubation time x, is based on exponential decay:

$$h\left(x;\underline{\theta}\right) = \theta_1\left(1 - e^{-\theta_2 x}\right).$$

Figure 1.1.e shows the data and the regression function.

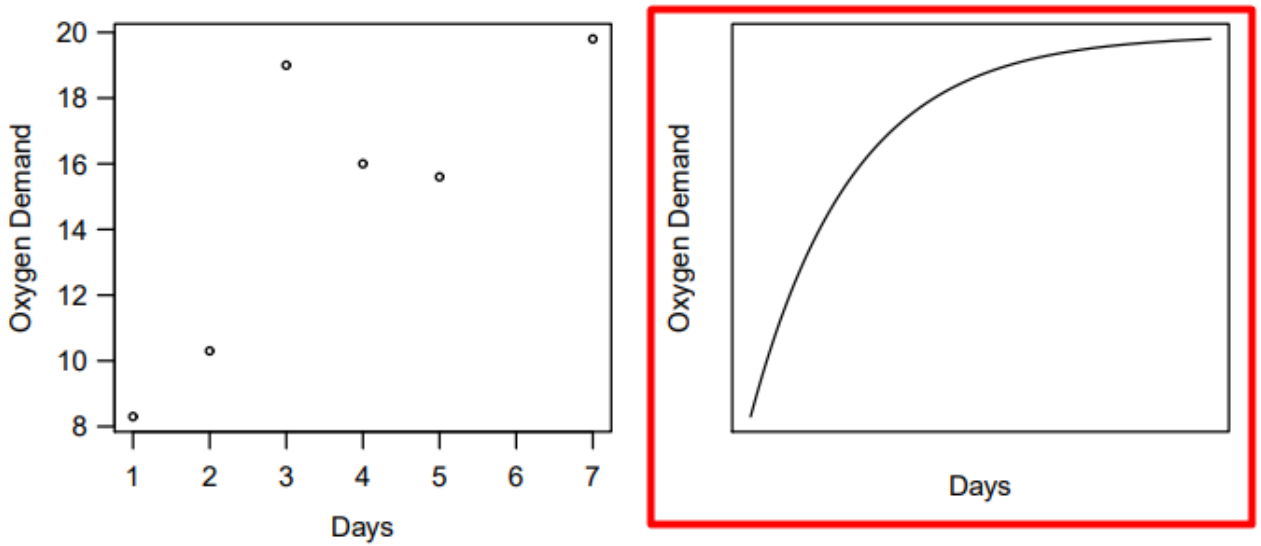
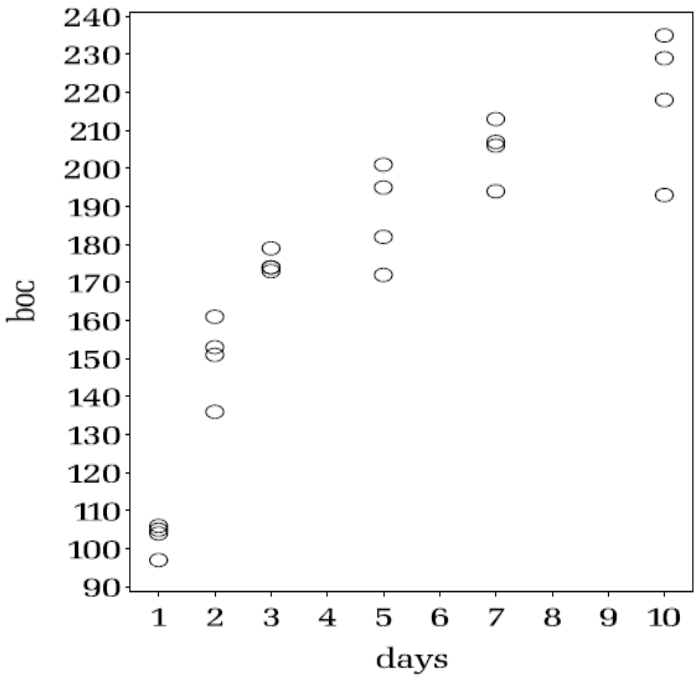


Figure 1.1.e.: Biochemical Oxygen Demand. (a) Data and (b) typical shape of the regression function.

Thoughts:

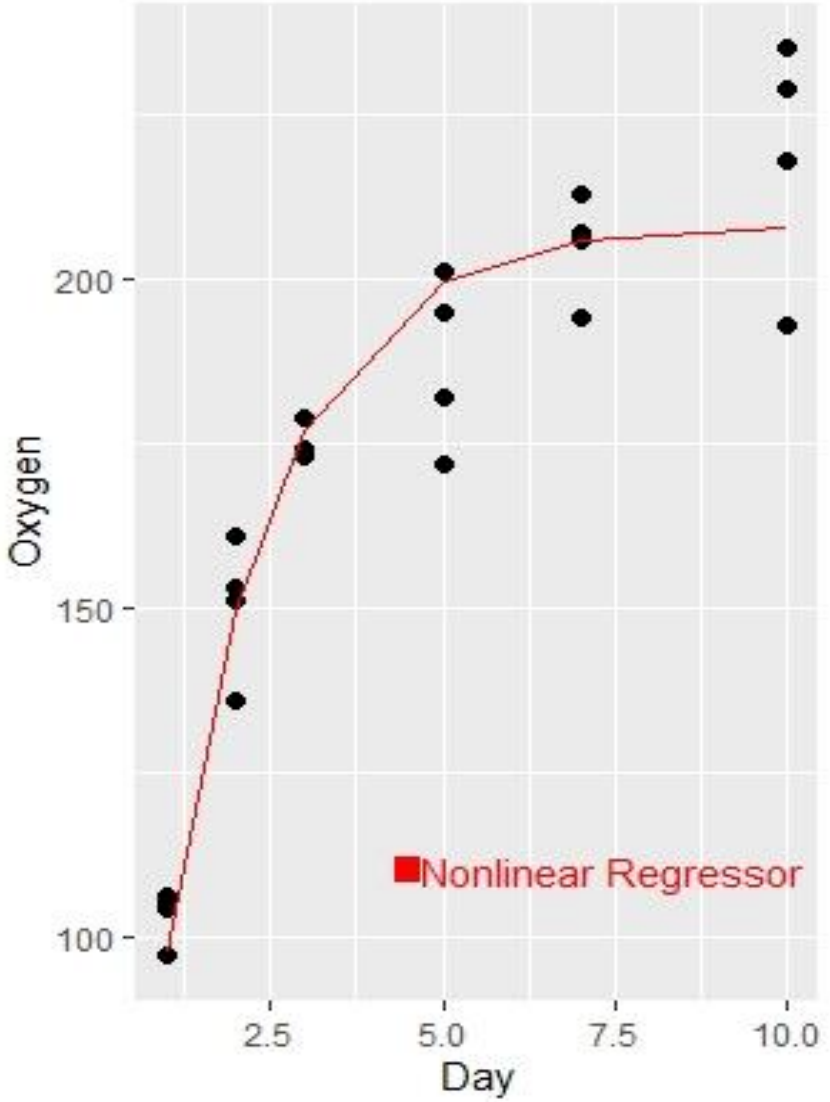
Comparing the data distribution and its possible curve from the course materials, I choose the nonlinear function shown by the left figures.

Data distribution of HW5



1. Find some nonlinear function with two parameters

$$h\langle x;\underline{\theta}\rangle = \theta_1\left(1-e^{-\theta_2 x}\right).$$




```
model_NL <- nls(y ~ theta1 * (1- exp(-theta2 * x)),
                data = df,
                start = list(theta1 = 300, theta2 = 10),
                trace = T
               )
summary(model_NL)
```

The upper function h is my selected model.
And I use the module nls to conduct the nonlinear regression.
The left figure shows the fitted curve of the nonlinear function h.
The following figure show the value of the two estimated parameters.

```
> summary(model)

Formula: y ~ theta1 * (1 - exp(-theta2 * x))

Parameters:
        Estimate Std. Error t value Pr(>|t|)
theta1 208.26843    4.09055   50.91  < 2e-16 ***
theta2   0.63232    0.04426   14.29  1.3e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.27 on 22 degrees of freedom

Number of iterations to convergence: 5
Achieved convergence tolerance: 7.5e-06
```

2. Do a hypothesis testing on the coefficients, both through linearization and through likelihood approach

Testing through Linearization

```
Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-39.552  -9.337   5.924  17.562  24.282

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  121.360      7.380  16.444 7.62e-14 ***
x             11.119      1.318   8.433 2.43e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 19.97 on 22 degrees of freedom
Multiple R-squared:  0.7638,    Adjusted R-squared:  0.753
F-statistic: 71.12 on 1 and 22 DF,  p-value: 2.432e-08
```

The p-values are quite small, so the intercept and the coefficient of x are significant.

Testing through likelihood:

```
> summary(model)

Formula: y ~ theta1 * (1 - exp(-theta2 * x))

Parameters:
       Estimate Std. Error t value Pr(>|t|)
theta1 208.26843    4.09055   50.91  < 2e-16 ***
theta2   0.63232    0.04426   14.29  1.3e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.27 on 22 degrees of freedom

Number of iterations to convergence: 5
Achieved convergence tolerance: 7.5e-06
```

The two p-values are quite small, so the two coefficients, theta1 and theta2, are significant.