Question 1:

We have a data set $\{(x_i, y_i) : 1 \leq i \leq n,\ x_i \in I\!R, y_i \in I\!R\}$ which is visualized in Figure 1 and which we want to describe by a nonlinear model:

$$Y_i = f(x_i; \theta) + \epsilon_i, \quad i = 1, \cdots, n$$

where the ϵ_i's are i.i.d. normally distributed random variables with $E\epsilon_i = 0$ and $E\epsilon_i^2 = \sigma^2$, unknown. Moreover, the regression function is given by

$$f(x; \theta) = \theta_1 + \theta_2 \frac{\exp(x - \theta_3)}{1 + \exp(x - \theta_3)}$$

with $\theta = (\theta_1, \theta_2, \theta_3)^T \in I\!R^3$.

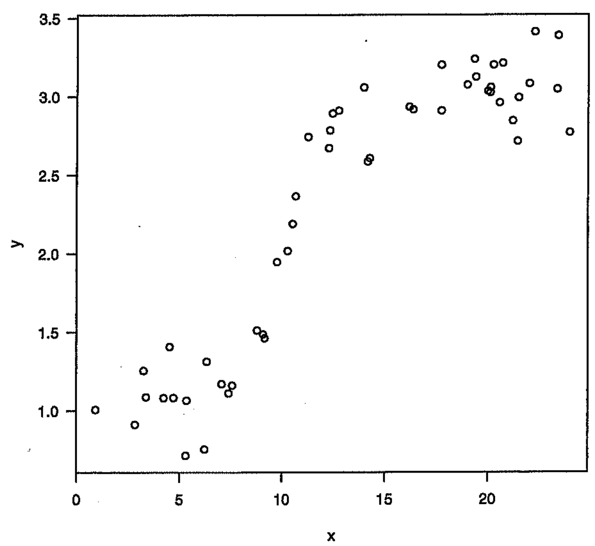


Figure 1: Scatter plot of the data

a) In order to estimate θ via maximum likelihood, one can employ the Gauss-Newton method. For this iterative numerical procedure, it is important to have reasonable starting values. Based on the data given in Figure 1, give a reasonable starting value for the vector θ and explain your choice.

b) Based on 50 observations, the parameter estimate resulting from the Gauss-Newton procedure is given by $\hat{\theta} = (1.05, 1.98, 10.15)^T$. The matrix $\hat{\sigma}^2 \left(\hat{V}^T\hat{V}\right)^{-1}$ that estimates the covariance matrix of the maximum likelihood estimator $\hat{\theta}$ looks the following:

$$\begin{pmatrix} 0.002610116 & -0.002509865 & 0.003893631 \\ -0.002509865 & 0.003715858 & -0.001865147 \\ 0.003893631 & -0.001865147 & 0.029469388 \end{pmatrix}$$

Construct an approximate 95%-confidence interval for θ_3 based on classical theory. If needed use $t_{47,0.95} = 1.6779$ and $t_{47,0.975} = 2.0117$ (You don't have to explicitely evaluate the result.).

c) Describe the procedure that could be followed for constructing a bootstrap confidence set for θ_3 of approximate level 95%.

Question 2:

The management of a paint-producing company has the choice between three types of paint-supplements needed to harden the paint. The first matter of interest is whether there is a difference in effect between the three supplements or not. For each supplement type, ten cans of paints are selected and treated with that type. The resulting thirty cans of paint are used to paint thirty pieces of iron. The paint is allowed to dry for two days, under controlled circumstances. Then the hardness of the paint is measured on a ratio scale for all thirty samples.

a) Formally describe the model to be used for the data obtained.

b) Describe the procedure (including hypothesis, test statistic and conclusion) of the relevant test.

Apart from the paint-hardening supplement, there is another factor that seems to be of importance for the hardness of the dried paint: whether the paint is water- or terpentine-based.

c) Formulate the model to be used here, together with three hypotheses that are usually of interest in these models.

Question 3:

One is interested in the behaviour of the probability of presence of a certain disease depending on the average daily volume of alcohol that is consumed.

a) Formulate a logistic regression model for this situation and give two reasons why this is a generalized linear model.

b) One can choose to include the explanatory variable in the model or not. How can this be tested?

The illness is usually a consequence of a too high concentration of a type of bacteria in the blood. A new measuring device allows the investigator to count the number of bacteria per unit volume in the blood, a quantity that is modelled as random variable with Poisson distribution.

c) Formulate the most natural GLM for this type of data, modelling the number of bacteria per unit volume of blood depending on the alcohol consumption described above.

Question 4:

Look at the time series plotted below in Figure 2, displaying monthly observations for 6 consecutive years.

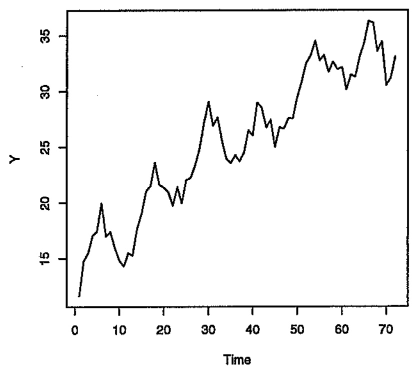
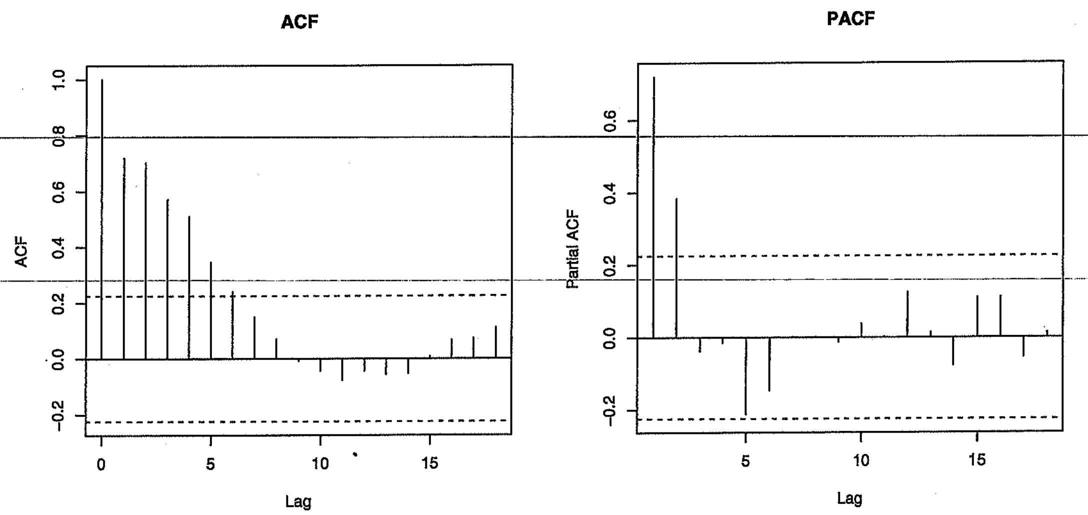


Figure 2: Time plot of the data

a) Give the definition of stationarity for a time series in general. Comment on what pattern you observe. Do you think the time series is stationary? Why or why not?

b) Formally describe one method to detrend and deseasonalize a time series.

The following plots show the ACF and PACF of a stationary time series:



c) Which type of time series model would you fit and why? Describe the model you have chosen formally.

d) Consider an MA process of order one. Derive the autocorrelation function for this model and explain how it can be estimated from the data.

e) What is the formal relationship between AR and MA models? What does it imply about the pattern of the autocorrelation function in AR processes?

Good luck with this exam!

Grading

1	2	3	4
a:2	a:3	a:3	a:3
b:2	b:3	b:3	b:3
c:3	c:3	c:2	c:2
			d:2
			e:2

Grade of written exam:

1+(number of points)/4

Final grade (provided the exam result is ≥ 5.5):

mean(weekly exercises)/3 + 2(grade of written exam)/3