Exam Statistical Models — Department of Mathematics
February 8, 2006 — *vrije* Universiteit

1. Describe for each of the following models how they differ from (or are specific instances of) the linear regression model.

   a) ANOVA

   b) Nonlinear regression

   c) Generalized linear regression

2. One is interested in the behaviour of a dichotomous response variable (that can only have the values 0 and 1) depending on two continuous explanatory variables that are measured on a ratio scale and two explanatory variables that are measured on a nominal scale with, respectively, two and four categories.

   a) Formulate a logistic regression model for this situation, giving attention to the way the nominal variables enter the model.

   b) Describe how the relevance of a certain explanatory variable in the model can be tested. Distinguish between the continuous and nominal explanatory variables.

3. We have a data set $\{(x_i, y_i) : 1 \le i \le n,\ x_i \in \mathbb{R}, y_i \in \mathbb{R}\}$ and want to describe this 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 $\theta = (\theta_1, \theta_2, \theta_3)^T \in \mathbb{R}^3$. Figure 1 shows a 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 motivate your choice using Figure 1.

   b) The matrix $\hat{\sigma}^2 \left(\hat{V}^T \hat{V}\right)^{-1}$ estimates the covariance matrix of the maximum likelihood estimator $\hat{\theta}$. For the given data set, this matrix is given by

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

   The parameter estimate resulting from the Gauss-Newton procedure is given by $\hat{\theta} = (1.05, 1.98, 10.15)^T$. Using this, construct a confidence interval for θ_3 of

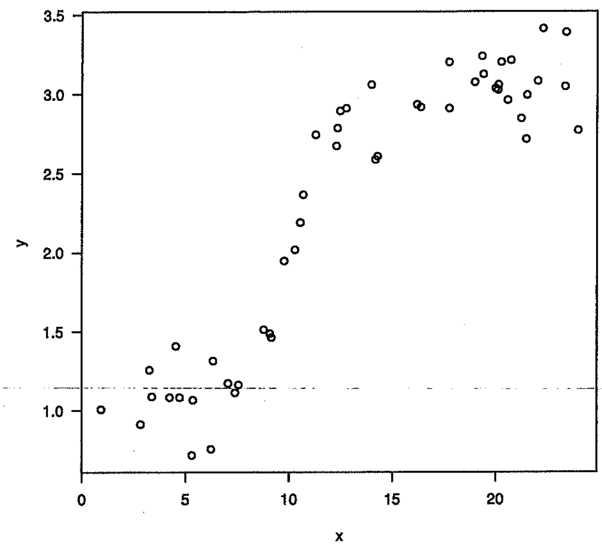

Figure 1: Scatter plot of the data

level approximately 95% based on classical theory. Also here, sums, ratios and powers of numbers may be used in the formula without explicitly evaluating these.

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

4. A Canadian dairy farm wants to know whether or not age and / or breed of cows have / has effect on the butterfat content of the milk produced. Random samples of 10 mature (five-years old and older) and 10 two-years old cows were taken from each of five breeds. For each of these cows, butterfat content of the produced milk was recorded. Assume that all other relevant influences are under control.

   a) Formally describe the model you want to use for the data obtained.

   b) Describe the procedure to test the relevant hypotheses (which, of course, you should also formulate).

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

   a) Comment on what pattern you observe. Do you think the time series is stationary?

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

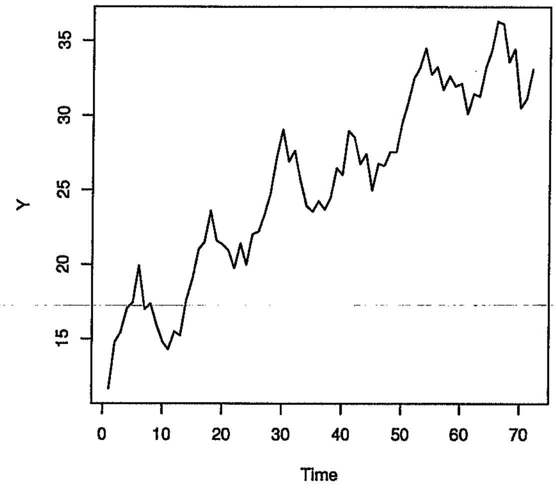

Figure 2: Time plot of the data

The ACF and PACF of a stationary time series are plotted in Figure 3 and Figure 4.

c) Which type of time series model would you fit and why? Describe the model you have chosen formally. Derive the autocorrelation function for this model and explain how it can be estimated from the data.

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

Grading

1	2	3	4	5
a:3 b:3 c:3	a:4 b:4	a:3 b:3 c:3	a:4 b:4	a:2 b:3 c:3 d:3

Grade of written exam: 1+(number of points)/5
Final grade:
mean(weakly exercises)/3 + 2(grade of written exam)/3, provided the exam result is $\geq$5.5

Good luck with this exam!

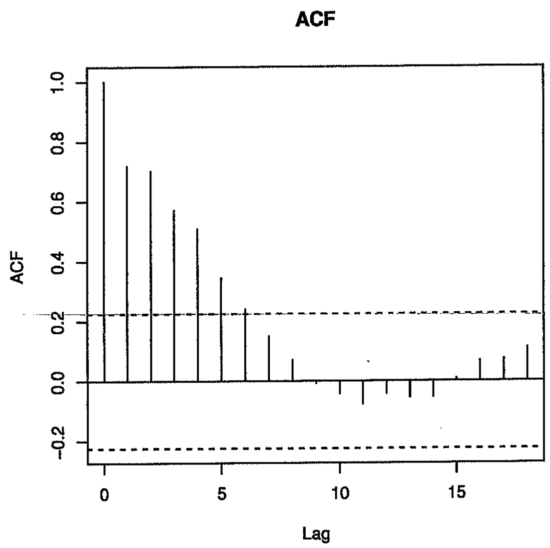

Figure 3: ACF of stationary time series

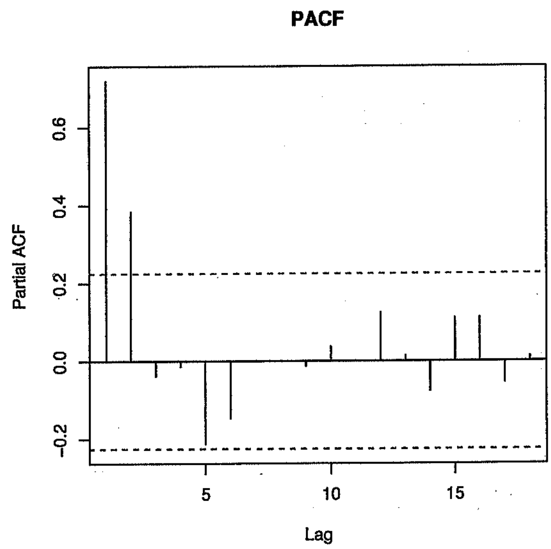

Figure 4: PACF of stationary time series