

BIOST 515: HW 1
Due: 1/15/2004 in lab

1. Use the SMSA data that was discussed in class to explore the relationship between mortality as the response and $\log_{10}(\text{NOx level})$ as the predictor.
 - (a) Show a scatterplot of NOx versus mortality and comment on any interesting trends you notice.
 - (b) Using least squares, fit the simple linear model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

where Y_i denotes the mortality for the SMSA and x_i denotes the log of the NOx levels.

- (c) Write out the model for a one unit change in x_i . How do you interpret β_0 and β_1 ? (Hint: remember $x_i = \log(\text{NOx}_i)$.)
- (d) Give parameter estimates and standard errors for β_0 and β_1 . Explain carefully how you would explain the values of these parameters to a non-statistician.
- (e) Give confidence intervals for these parameters, carefully stating any assumptions that you have made.
- (f) Obtain a point estimate for the variance, σ^2 .
- (g) Add the fitted line with confidence intervals to the scatterplot.
- (h) Repeat a-g with the data set subsetted to those SMSAs with more than 20 inches of rain fall per year.
- (i) Comment on the differences between the fits of the two models. You may want to look into the subset argument to `lm()`.

The SMSA data set that was discussed in class is available on the class website.

2. Pharmacokinetics is the study of the time course of a drug and its metabolites after its introduction in the body. The concentration of a drug at time x , $c(x)$, is often modeled as

$$c(x) = \left(\frac{D}{V}\right) \exp(-Kx).$$

This deterministic model predicts the concentrations of the drug in the body as a function of time. It depend on two unknown parameters, the elimination rate constant K and the volume V . D is the amount of drug administered.

In order to see if an experiment would provide sufficient information, It is decided to simulate a set of data for a patient from a particular set of sampling times. Suppose the patient is to be administered a single

dose of size $D=10$ mg. It is believed that the volume of a patient's blood compartment is $V=5$ liters and that the elimination rate of the patient is $K=0.7$.

The concentrations (in mg/L) for this patient as a function of time, x , are then given by

$$2.0 \times \exp(-0.7x).$$

The logarithms of the concentrations are given by the linear model

$$\log 2.0 - 0.7x. \tag{1}$$

Suppose that the log of observed concentration at time x is given by this equation with the addition of an error component that arises from the $N(0, 0.2^2)$ distribution.

- (a) Simulate log concentrations for such an individual where it is assumed that the concentration is measured at the following times (in hours):

(0.5, 1.0, 2.0, 3.0, 4.0, 6.0, 8.0, 10.0, 12.0, 24.0).

To perform the simulation, you must first calculate the expected concentration at the given times using equation (1). Then add error terms that are simulated, independently, from the normal distribution. Denote the observed log concentration / time point pairs by (y_i, x_i) .

- (b) Assuming a relationship of the form

$$E[Y|x] = \beta_0 + \beta_1 x$$

find, for the data you have simulated, the least squares estimates of β_0 and β_1 .

- (c) Obtain confidence intervals for β_0 and β_1 and compare them with the values that were used to generate the data.
- (d) Plot the data with the least squares regression line.
- (e) Consider a hypothetical patient who provided the simulated data. What dose would you give this patient in order to obtain a concentration of 6mg/L at 2 hours?

You will need to use the R function `rnorm()` to simulate the error terms from the normal distribution.

3. KKMN 5.12. Do not rely on the SAS output given with the problem. Where the problem says "sketch" do not hand sketch, but plot on a computer. For part (e), it is not necessary to sketch your idea of the true relationship, you may state it in words.