Exercises 1

1. (This is a repetition of the two-samples t-test from Rice, Section 11.2.1.)
   A study was done to compare the performances of engine bearings made of different compounds. Ten bearings of each type were tested. The following table gives the times until failure (in units of millions of cycles):

Type I	Type II
3.03	3.19
5.53	4.26
5.60	4.47
9.30	4.53
9.92	4.67
12.51	4.69
12.95	12.78
15.21	6.79
16.04	9.37
16.84	12.75

   (a) Write down an appropriate linear model (using the general form of linear models) for these data and use the model (with normal errors) to test the null hypothesis that that the two expected values of the two types are equal against a two-sided alternative. State all the model assumptions that are needed. (Note that this leads to the t-test in Rice Section 11.2.)

   (b) Discuss whether the model asumptions are reasonable for these data, using suitable data visualization.

   You may apply the Mann-Whitney test to the data as well if you like.

2. Invent examples of data analytic problems, one each for the three major aims of modelling:

   (a) prediction,

   (b) explanation,

   (c) causal inference.

   By "invent" it is meant that you describe the variables, what scientists are interested in, and the background of the situation as far as necessary to understand to which of the three aims you classify the example. You do not need to make up the datasets.

3. *Read Rice section 14.3:* this shows how to derive the normal equations for least squares estimation of the parameters in a multiple linear regression model and how these equations are written in matrix form (as in Chapter 2 of the lecture notes). Check that you can obtain the normal equations shown in Rice and understand how they can be summarised in matrix form by carrying out the following for the case of two explanatory variables.

   (a) Express the sum of squares of the errors in terms of the three unknown parameters $\beta_0, \beta_1, \beta_2$, and then obtain the normal equations by differentiation with respect to each parameter.

(b) Write down the matrix $\mathbf{X}$ and the vectors $\boldsymbol{\beta}$ and $\mathbf{Y}$ and verify that the equations obtained in (i) are of the form $\mathbf{X}^{\mathrm{T}}\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^{\mathrm{T}}\mathbf{Y}$.

4. In your preparatory reading from Rice's book for the Foundation Course, you should have revised the bivariate normal distribution. Refer to the relevant sections in Rice's book:
section 3.3 example F for pdf,
section 3.5.2 example C for conditional distribution,
section 4.3 example F for covariance and correlation,
section 4.4.1 example B for conditional expectation and variance.

The pdf of the (k-dimensional) multivariate normal distribution can alternatively be given in terms of the mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, as

$$f(\mathbf{y}) = (2\pi)^{-\frac{k}{2}}(\det\Sigma)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\}$$

In the bivariate case, show that this formula for the pdf gives the pdf shown in Rice.