COMPUTER EXERCISE 2: MULTIPLE LINEAR REGRESSION

A list of useful R-commands for multiple linear regression is found on the course home page

```
http://www.maths.lth.se/matstat/kurser/masm22/lab1_vt19_useful.pdf
http://www.maths.lth.se/matstat/kurser/masm22/lab2_vt19_useful.pdf
```

See also the lecture R-files for examples.

Exercise 2: Sleep in Mammals

A study on sleep in mammals collected, among other things, the body weight, the average time spent sleeping during a day (24 hours), and a danger index for 57 species of mammals. Of interest is to predict the time a species spends sleeping, using the body weight and the danger index.

Variable	Description
Species	Name of the species
TotalSleep	total sleep (hrs/day)
BodyWt	body weight (kg)
Danger	overall danger index (1–3)
	1 = low danger (from other animals);
	2 = medium danger (from other animals);
	3 = high danger (from other animals)

The data for this exercise is available as a tab-separated txt-file, `sleep.txt`, on the course web page. Download and save it to your R working directory and then read it into R. Since the danger index is a categorical variable, we should also turn it into a factor in R:

```
sleep <- read.delim("sleep.txt")
sleep$Danger <- factor(sleep$Danger, levels = c(1, 2, 3),
                       labels = c("low", "medium", "high"))
```

Danger

(a) Let's start with some exploratory analysis. Make a frequency table of the Danger variable. How many observations are there in each category?

(b) Calculate the average daily sleep in each of the three Danger categories. Then plot TotalSleep versus Danger (this should automagically become a boxplot) and compare with the calculated averages.

(c) Fit the linear model $\text{TotalSleep}_i = \beta_0 + \beta_1 \text{Danger}_{\text{medium},i} + \beta_2 \text{Danger}_{\text{high},i} + \varepsilon_i$, using the Danger variable and letting R choose the default reference category. Use the model to estimate the average amount of sleep for animals in each of the danger categories. Compare the result with the category averages you calculated above and make sure you understand how the β-parameters relate to the category averages.

(d) Use a (global) F-test to test whether we need the Danger variable.

(e) Calculate 95 % confidence intervals for the three β-parameters.

Danger and body weight

(f) Obviously, the amount mammals sleep varies with the amount of danger. We now want to see if body weight affects the sleep. Plot total sleep vs body weight. Does a linear relationship seem reasonable? Redraw the plot using the log of the body weight instead. Does this seem more reasonable? Why might this be more reasonable?

(g) First fit a model using log(BodyWt), but not Danger. Then fit a new model adding Danger as well. Use a (partial) F-test to test whether we need the Danger variable now. Also test whether we need log(BodyWt) when we have Danger.

(h) Calculate 95 % confidence intervals for the four β-parameters. Are there any interesting differences for any of the parameters, compared to (e)? Why might that be? It might be helpful to plot log(BodyWt) against Danger.

(i) Check that the residuals follow the assumptions by examining the following plots:(1) Q-Q-plot, (2) residuals vs predicted values, (3) residuals vs log(BodyWt), (4) residuals vs Danger. Anything that might indicate a problem?

Predictions

(j) We have discovered a new species that weighs 30 kg and lives in medium Danger. Calculate a 95 % prediction interval that would be expected to contain the total sleep of this new species.

(k) Homo Sapiens is a mammal with body weight 62 kg and low Danger. Estimate its expected total sleep together with a 95 % confidence interval.

(l) The Marmot and the Vervet both weigh 4 kg but the Marmot lives in low danger and the Vervet in high danger. How many hours less would the Vervet be expected to sleep, compared to the Marmot? Write down the theoretical value, estimate it and calculate a 95 % confidence interval.

(m) How many hours longer would the small Marmot be expected to sleep, compared to the large Homo Sapiens? Write down the theoretical value, estimate it and calculate a 95 % confidence interval.

(n) How many hours longer (or shorter) would the small, high danger Vervet be expected to sleep, compared to the large, low danger Homo Sapiens? Write down the theoretical value, estimate it and calculate a 95 % confidence interval.

Hint: since $V(\mathbf{a}\hat{\beta}) = \mathbf{a}\,Var(\hat{\beta})\mathbf{a}'$, you will need the covariance matrix for the β-estimates.