TMA4255 Applied Statistics
Exercise 3

Problem 1 - Simple linear regression (theory)

In simple linear regression

$$E(Y_i) = \beta_0 + \beta_1 x_i,\ i = 1, ..., n$$

where n is the number of observations.

a) Let B_0 and B_1 be the least squares estimators for β_0 and β_1 given in Section 11.3 in the book. Find expressions for the variances of B_0 and B_1 and their covariance.

Hint: You may use, without proving it, that $\bar{Y}$ and B_1 are independent.

b) Derive the expressions given in Section 11.6 in the book for a $(1-\alpha)\cdot 100\%$ confidence interval for the expected response when $x = x_0$ and a $(1-\alpha)\cdot 100\%$ prediction interval for Y_0 in the same point. Which value of x gives the shortest interval?

Problem 2

Here we will look at an experiment conducted by Forbes to study the relationship between the boiling temperature of water (in Fahrenheit) and the pressure in the atmosphere (in inches of Mercury).

Forbes' data from 1857 give the boiling temperature and the corresponding barometric pressure at 17 locations in the Alps and in Scotland.

Temp. (T_i)	194.5	194.3	197.9	198.4	199.4	199.9	200.9	201.1	201.4
Pressure (p_i)	20.79	20.79	22.40	22.67	23.15	23.35	23.89	23.99	24.02
Temp. (T_i)	201.3	203.6	204.6	209.5	208.6	210.7	211.9	212.2	
Pressure (p_i)	24.01	25.14	26.57	28.49	27.76	29.04	29.88	30.06	

Throughout this exercise we assume treat temperature (boiling point) as the response variable and pressure as the dependent variable.

a) Look at the relationship between pressure and temperature by constructing a scatter plot. Does it look linear?

If we assume that $\text{E}(T_i) = \alpha_0 + \alpha_1 p_i$, we may fit a simple linear model with temperature as response and pressure as covariate. Plot the standardized residuals against the pressure. In what way does this plot tell you that the model does not fit the data very well?

MINITAB: Data are found at the Exercise tab of the course www-page. Scatter plot is found under the `Plots` meny and then `Scatter plot`. If the temperature is put in C1 and the pressure in C2 the simple linear regression can be performed by:

Stat → Regression → Regression

Response:	C1 (T)
Predictors:	C2 (p)
Storage:	x Fits
	x Residuals (not really necessary)
Graphs:	Standardized residuals vs. the variables: C2 (p)

R: Data are already available in R in `library(MASS)` (which probably already is installed). Object `forbes` contains the data. Scatterplot by `plot(forbes$pres,forbes$bp)`. Linear model by `fit=lm(bp~pres,data=forbes)` and `summary(fit)`. Plotting the standardized residuals vs. pressure by `plot(forbes$pres,rstandard(fit))`.

b) Forbes believed that the boiling point (temperature) was a linear function of the logarithm of the pressure. Transform the pressure values such that $x_i = 100\ln(p_i)$. Does the model $\mathrm{E}(T_i) = \beta_0 + \beta_1 x_i$ fit the data better than the model in a)?

MINITAB: Calc-Calculator. Choose "Natural log", and NOT ln gamma (which is just a linear transformation). **c)** We want to test $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$.

Write down the expression for s^2 and the t-statistic, and compute using the data. Does the t-test reject the null hypothesis if we use a significance level of 1%?

d) How would the test in c) be performed using the analysis of variance table? Show that these two tests are the same.

MINITAB: Available from the linear regression output.
R: `anova(fit)`

e) How would we test $H_0 : \beta_1 = 1$ versus $H_1 : \beta \neq 1$? Use 1% significance level.

f) Construct a 99% confidence interval for β_1 and explain how one from this get to the conclusion in e).

R: `confint(fit,level=0.99)`